Why Do Diplomas Pay?

An Expanded Mincerian Framework Applied to Mexico.

Aashish Mehta
Asian Development Bank*
Manila, The Philippines

Hector J. Villarreal
Department of Economics & School of Public Policy
ITESM Campus – Monterrey, Mexico.

(Words – 8,600)

Abstract:

We compare four explanations for the value of diplomas. We first note that each explanation has implications for unemployment and wage variation amongst graduates. Next, we test for these conditions using a novel econometric framework, exploiting idiosyncrasies in Mexican labor market and educational institutions. Premiums appear to result from diplomas tied to jobs with downwards rigid wages – a non-standard explanation. The standard explanations, including signaling, are not supported by the data. Our results illuminate how labor markets segmented by diplomas clear. This depends upon the nature of the labor market rigidities exactly as predicted by neoclassical theory.

Notice: The views presented in this paper are solely those of the authors and do not represent those of the Asian Development Bank.

Acknowledgements: We would like to thank Jean-Paul Chavas, Ian Coxhead, Michael Carter, Bill Provencher, Peter Norman, Jorge Aguero, Brad Barham, Bonnie Palifka, Brian Gould, Diana Weinhold, Jeremy Foltz and participants at the Development Economics workshop at the University of Wisconsin-Madison for their comments. All errors are our own.

I. Introduction:

Mincer's (1974) simple specification for a log-wage equation has been extended in a number of ways to examine patterns in the relationship between schooling and earnings. The percentage increase in wages statistically attributable to a particular component schooling is referred to as the Mincerian rate of return to that component. Hungerford and Solon's (1987) specification has allowed many authors to ask whether completing the last year of a particular level of schooling confers an above-average proportional wage increase. Such a completion premium is referred to as a *Sheepskin Effect*. Similarly, several authors have also examined the relative Mincerian returns to different schooling streams, such as formal and vocational training.

A long and unresolved debate has ensued over how to interpret the results to these studies. Broadly, the more common explanations can be usefully caste into three categories. Some authors, building on the work of Arrow (1973) and Spence ( ), ascribe their results to the productivity signals generated by different diplomas. Others refer to two lines of reasoning traceable to Chiswick (1973). One holds that the wage premium on diplomas could simply reflect above average productivity increases conferred by the completion of diploma years of schooling. The other relates returns to schooling to self selection, due to stronger students, who will also be better workers, graduating in higher numbers from harder programs. This paper develops a fourth class of explanation – that these patterns are driven by institutional rigidities in markets for labor and education, and examines the evidence in favor of these four interpretations.

The principle underlying the investigation is simple. Each theory, taken individually, implies patterns in the returns to different diplomas and courses of study. We also note that each theory often has implications regarding the employment propensity, length of work-week and the variance of wages of workers with different educational backgrounds. We begin by elucidating why these theories imply these patterns.

We then propose and estimate a three-equation maximum likelihood specification in which employment, hours worked and log-wages, as well as their joint variance matrix, are conditioned on the standard Mincerian variables. Because wages are only observed for the employed, the employment equation takes a probit form, and wage and hours equations are only estimated for the employed. This framework enables us to identify discontinuities at diploma years in the correspondence between education, on one hand, and log-wages, employment propensity, hours worked and the variance of wages, on the other (*wage, employment, hours* and *wage-variance* *sheepskin effects* – if you will). The tests for employment and hours effects of diplomas permit us to examine how labor markets, segmented by diploma-driven discontinuities in the wage-schooling schedule, clear. Comparison of these patterns across different schooling streams similarly sheds light on how markets for different skills clear.

The policy implications of each of the theories being contrasted are profoundly divergent. If signaling is found to drive the returns to diplomas, then it would be useful to develop (if possible) more efficient signaling mechanisms (standardized testing and the like) which could replace components of the educational experience which do not enhance productivity. In contrast, neither of Chiswick's arguments engenders any inefficiency. Evidence in favor of either, individually, therefore carries no rationale for policy change for efficiency reasons. However, as self-selection and variations in productivity improvements conferred by different schools may be driven by schools' resource constraints, such evidence might embolden those who emphasize the distribution of educational expenditures. Finally, finding that institutional rigidities are the real culprits should promote a robust debate regarding the objective, performance and side-effects of these institutional arrangements (minimum wage laws, certification requirements, and the like).

Detailed empirical work to distinguish between the different explanations is therefore extremely valuable. We find that the data from Mexico are strongly consistent with the fourth, institutional explanation, and inconsistent with each of the other theories taken on their own. Despite the heuristic appeal of the institutional rigidity explanation, we have not found entries in the literature that present it when interpreting their results.[1]

The remainder of the paper is structured as follows. In the next section, we briefly review the literature on measuring the returns to schooling, and highlight insights from this literature that are crucial to the current exercise. Section III presents the four theories of diploma premiums, and their divergent implications for sheepskins in the first and second moments of the earnings distribution, hours worked and in employment propensity. It also details a fortuitous alignment of institutional circumstances in Mexico which allows us to test implications of our job-specific diploma explanation of sheepskin effects. Section IV introduces the data and provides a description of the Mexican education system. In section V, we present the econometric specification and formalize our hypothesis tests. Section VI presents and interprets the model estimates and hypothesis test results. Section VII concludes. One

[1] For a classic example of a paper that examines the three standard explanations, but not the fourth, see Schady (2003)

appendix derives the likelihood function used, and another formalizes the argument for a second-moment sheepskin effects under Arrow's screening model.

II. Measuring the Returns to Schooling

Studies seeking measures of the returns to schooling fall into two categories. The first group, summarized most elegantly by Card (2001), carries Chiswick's intuition regarding self-selection into the econometric domain. These studies argue that because the decision to acquire additional schooling is not random, and because the private marginal benefits and costs of schooling are probably negatively correlated across workers, workers most likely to obtain high returns to schooling acquire more of it. Hence, OLS estimates of rates of return to schooling and diplomas overestimate the true treatment effects. These studies employ a two stage instrumental variables (IV) procedure, often utilizing variations in institutional arrangements to instrument for schooling in order to purge the earnings equation of individual effects. Unfortunately, the estimated rates of return calculated in these studies are even higher than OLS estimates[2], perhaps reflecting the paucity of suitable instruments.

The second class of studies tends to ignore the endogeneity bias problem, focusing instead on contrasting returns to different elements of the schooling experience (for example – diploma vs. non-diploma years, private vs. public schooling etc.). While our interest in sheepskin effects necessitates that our study originate in this second tradition[3], we draw on three key insights from the two-stage literature. First, conventional estimates of returns to schooling, and especially, diplomas, are likely to be inflated by selection into treatments. This renders tests confirming the absence of sheepskin effects more powerful. Second, workers differ, so that documenting the heteroskedastic structure of education and labor market outcomes is crucial to characterizing equilibrium behavior. And finally, institutional arrangements play a critical role in shaping schooling decisions and the returns to education.

Consideration of the last point led us to consider how the consequences of diploma holding would vary with institutional conditions, and thus, to our institutional theory of diploma effects, presented in the next section.

[2] See, for example, Alderman et. al. or the results summarized by Card (2001).
[3] It is difficult to see how a sufficient number of instruments could be found for the different types and levels of schooling we are interested in.

Picking up on the second point raised by the IV literature on returns to endogenous education, we worry about the economic interpretation of heteroskedasticity, as explained in the next section. We also confront the econometric consequences of heteroskedasticity. Specifically, if the distribution of wages, conditional on schooling, is heteroskedastic, then restrictions on the simple Mincerian coefficients are insufficient for testing hypotheses regarding actual (not logged) wages. We demonstrate how to include restrictions on the conditional variance of log-wages, and show that failure to do so leads to empirically relevant test misspecifications.

The first point raised by the IV literature implies that if we take education to be exogenous, but still fail to find a wage sheepskin, the probability that this is a false negative is even lower than our test statistics suggest.

III. Testable Implications of Explanations for the Effects of Diplomas

The following section sketches out the testable implications of each explanation for the effects of diplomas, particularly as they relate to sheepskin effects. Table 1 summarizes these implications, explained heuristically below.[4]

Layard and Psacharopoulos (1974) argued that positive sheepskins in wages are necessary implications of Arrow's (1973) *screening theory* [5]. This theory remains the most cited interpretation for sheepskin effects in the empirical literature to date.

The screening argument states that diploma receipt conveys positive information to employers regarding potential employees that could not be inferred in its absence. This argument for a diploma effect requires two presumptions. First, that the information inferable from diploma receipt is greater than what is inferable from the successful completion of a non-diploma school-year. Second, it requires that this information is singularly helpful to all graduates. Picking up on this second assumption, we argue that if a diploma also provides additional data – grades, for example - which permit wage offers to potential employees roughly in accordance with their skills, a sheepskin effect will become evident in variance of earnings. This is because diplomas demonstrating poor academic performance will carry small wage premiums, while those indicative of greater talent will carry larger ones. If diplomas carry more than a binary signal, therefore, the conditions necessary for a first order sheepskin due

[4] The heuristic approach has been chosen for reasons of brevity. Each implication can be theoretically derived.
[5] Riley (1979) argues that screening theory, in certain circumstances, does not require a sheepskin effect, so that the requirement should be properly softened to a non-negative wage sheepskins.

to screening are likely to result in a second moment sheepskin. [6] We formalize this argument in the style of Arrow (1973) in Appendix B. We also note that the screening hypothesis does not, in general, have anything to say about the effects of diplomas on a worker's propensity to be employed or the length of their work-week, as this outcome depends critically on how labor demand and supply modeled.

Chiswick's (1973) first explanation of sheepskins holds that worker productivity may not increase smoothly with years of schooling. Following Hungerford and Solon (1987), several studies have therefore allowed for non-linearities in the Mincerian equation at all levels of the school system.[7] Some of these find non-linearities at non-diploma years, and argue that large wage increases at given years of schooling may simply reflect threshold effects in the productivity-schooling schedule. Exactly what drives these threshold effects is left tantalizingly open.

Clearly, in theory, non-linear productivity increases can give rise to a first moment sheepskin effect. However, we see no compelling reason why such an explanation should result in wage-variance sheepskin, nor one why it shouldn't. In contrast, because this explanation does not involve a violation of Walras' law, the markets for degreed and non-degreed workers should clear. Because we have no voluntarily unemployed people in the sample, this theory, without assuming distortions, implies that diplomas will have zero empirical impact on observed employment propensity. On the other hand, as we cannot assess whether observed variations in hours worked are voluntary or reflect underemployment, the theory is ambiguous with respect to the effect of diplomas on hours worked.

Chiswick's (1973) second argument holds that if weaker students do not graduate, and if academic ability relates positively with productivity in the work-force, then sheepskin effects will result simply from self selection. This argument is distinct from the screening explanation of sheepskins, because it would hold even if employers could observe workers' productivities perfectly.

This explanation of sheepskins should result in first order wage sheepskins. It can, but need not, result in second order sheepskins. This would depend upon the precise statistical relationships between academic ability, workforce productivity and the utility cost of schooling. However, as the assumptions underlying a pure self-selection explanation of sheepskin effects do not violate the Walrasian assumptions, the market for degreed and non-

[6]The results become even starker if students are uncertain about their likely diploma quality. Those who perform unexpectedly poorly could even receive a negative diploma effect. If there were enough such disappointments, no first moment sheepskin would exist on average, while second moment sheepskin effects would be significant.

[7] For example: Arabsheibani and Manfor (2001) and Patrinos (1996).

degreed workers should still clear, and no employment sheepskin is expected. Finally, we note that in general, the impact of diplomas on hours worked is ambiguous in general because our hours data. However, if disutility from schooling and labor effort are positively correlated, graduates are likely to work longer hours. Finally, as an additional reality check, we note that a convincing self-selection explanation of a wage sheepskin also requires a reason why finishing a diploma year should be harder than a non-diploma year.

We now propose a fourth, institutional explanation for the impact of diplomas. Our job-specific diploma hypothesis (JSDH) holds that returns to diplomas might arise if particular diplomas are necessary for obtaining particular jobs, and the wages paid by these jobs are shielded from competition. To see this, suppose that particular jobs are tied to having a certain diploma. If minimum wage laws or collective bargaining are prevalent, or if the number of degrees to be granted is limited, the wage premium to diploma holders could persist. On the other hand, if such rigidities were uncommon, more students would be motivated to obtain this diploma, and the premium it carried would be eroded.

This hypothesis has different testable implications for employment, depending on what prevents diploma holders' wages from falling. Specifically, if this role is filled by minimum wage laws or across-the-board collective bargaining, then the market for degreed workers should clear through unemployment, implying a negative employment effect. The mathematics and logic of this argument parallel Harris and Todaro (1970) precisely. On the other hand, if wages paid to diploma holders cannot fall because of limitations on the number of diplomas issued, then full employment is expected in the degreed sector, implying a non-negative employment effect.

Now, diploma holders, under the job-specific diploma hypothesis, may obtain all the jobs that are open to non-diploma holders, as well as those requiring diplomas. It follows that diploma holders could obtain a wider range of wages, and we might see a sudden increase in the variance of earnings at diploma years.

Exactly how the hours worked by diploma holders will compare to those of other workers under the JSDH depends upon institutional factors in the two labor markets. Consider the following likely examples – first suppose that a particular diploma allows workers who would otherwise be confined to the informal sector, to secure formal sector jobs. If formal sector employers are required to bear fixed costs to hire a worker, they are likely to use a smaller work force more intensively. Then the diploma will be associated with a longer work week. Alternately, if a diploma qualifies workers to enter a unionized sector, it will be associated with shorter hours.

We distinguish between recipients of teaching certificates, technical high-school degrees and conventional high-school degrees in order to capture this story. We are fortunate to have such a comparison available for the following reasons: conventional high school degrees are not tied to particular professions, while technical high school degrees and teaching certification are required for specific jobs; and teachers' earnings are protected due to their position as unionized public sector employees,[8] 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. Any wage premium on technical, relative to formal degrees, should be eroded by increased competition for jobs requiring them, while the premium on teaching degrees cannot be eroded by economic competition. Further, because the number of teaching degrees granted in Mexico is limited, largely through union influenced restrictions on the number of students admitted to teaching programs, we expect low unemployment amongst teachers. Finally, holders of teaching degrees, being mostly unionized, will work shorter work weeks.

IV. The Empirical Environment:

The Mexican education system is a mixture of public and private institutions. The public institutions depend on federal, state or municipal governments for funding. Even though many children attend kindergarten, it is not an official prerequisite for admission to most primary schools.[9] Usually, twelve years of formal education are completed prior to college: six of primary school, three of junior-high and three of high school. College typically takes five years to complete, although the duration does vary.

Parallel to the formal education track, analogous levels of technical education exist, which provide a similar curriculum to the formal school system, complemented by vocational training. We limit our analysis to technical high schools, dropping the few 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, although we are unable to distinguish its recipients from other college graduates.

[8] 139 out of 173 identifiable teachers in our sample are union members.

[9] This may vary according to states or the kind of school. Many private schools do require some preprimary education.

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.

We pare down our sample using criteria that are standard in this literature, restricting our sample to non-students between the ages of 16 (the legal working age) and 65. We include employees and unemployed members of the work-force. [10] 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.

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 Behrman and Deolalikar (1991) and Jaeger and Page (1996), point out, this is important because imputing completion from temporal data can bias results.

V. The Model and Hypothesis Tests:

The Mincerian Equation

A Mincerian equation often takes the following form:

(1) $$y = \ln w = \delta^0 + \delta^E E + \delta^{E2} E^2 + \sum_{l=p,j,h,c} \left(\delta^{sl} s_l + \delta^{Dl} D_l \right),$$

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 (p̲rimary, j̲unior-high, h̲igh-school and c̲ollege). 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. The growth rate of wages with years of experience and of schooling at level l are $\delta^E + 2\delta^{E2}E$ and δ^{sl} respectively. Similarly, $\exp\left(\delta^{Dl}\right) - 1$ 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, δ^0 is permitted to vary with personal characteristics. Notice that a specification that "corrects"

[10] 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. Hence, it is advisable, and standard, to discard observations of self-employed workers.

for such personal characteristics through δ^0 still imposes constant returns to education and experience with respect to these characteristics.

We generalize this specification before embedding it in a larger three-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)
$$y = \ln w = \delta^0 + \delta^E E + \delta^{E2} E^2 + \sum_{l=p,j,h} \left(\delta^{sl} s_l + \delta^{Dl} D_l \right) + D_h \left(\delta^{sc}_{PF} s_c + \delta^{Dc}_{PF} D_c \right) + D_T \left(\delta_{DT} + \delta^{sc}_{PT} s_c + \delta^{Dc}_{PT} D_c \right) + \delta^{Teach} D_{Teach},$$

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*. Our sample excludes those who began, but 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 $\exp(\delta^{DT}) - 1$, and annual average rates of return to the degree of $\delta^{DT}/3$. 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)
$$\begin{bmatrix} z_i^* \\ h_i^* \\ y_i^* \end{bmatrix} = \begin{bmatrix} \beta x_{zi} \\ \gamma x_{hi} \\ \delta x_{yi} \end{bmatrix} + \begin{bmatrix} \varepsilon_{zi} \\ \varepsilon_{hi} \\ \varepsilon_{yi} \end{bmatrix}; \; \varepsilon_i = \begin{bmatrix} \varepsilon_{zi} \\ \varepsilon_{hi} \\ \varepsilon_{yi} \end{bmatrix} \sim N(0, \Sigma_i); \; \Sigma_i = \begin{bmatrix} 1 & \rho_{1i}\theta_i & \rho_{2i}\sigma_i \\ \rho_{1i}\theta_i & \theta_i^{\,2} & \rho_{3i}\theta_i\sigma_i \\ \rho_{2i}\sigma_i & \rho_{3i}\theta_i\sigma_i & \sigma_i^{\,2} \end{bmatrix};$$

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

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

$y_i = y_i^*$ if $z_i^* \geq 0$ 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*. θ_i and σ_i are the standard deviations of the "unexplained" components of the hours and logged earnings potentials respectively. Each of the ρ_{ki} is a correlation coefficient between unobservable components.

The allowance for heteroskedasticity is implemented via the Cholesky decomposition such that:

(4a) $$\varepsilon_i = A_i u_i \,;\; u_i \sim N(0, I_3);\; A_i = \begin{bmatrix} 1 & 0 & 0 \\ a_{3i} & a_{1i} & 0 \\ a_{4i} & a_{5i} & a_{2i} \end{bmatrix};\; a_{ji} = \alpha_j^0 + \boldsymbol{\alpha}_{\mathbf{j}}'\mathbf{x}_{\mathbf{ji}}.\; j = 1,..,5.;$$

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

(4b) $$\Sigma_i = V(\varepsilon_i) = A_i A_i' = \begin{bmatrix} 1 & a_{3i} & a_{4i} \\ a_{3i} & a_{1i}^2 + a_{3i}^2 & a_{3i}a_{4i} + a_{1i}a_{5i} \\ a_{4i} & a_{3i}a_{4i} + a_{1i}a_{5i} & a_{2i}^2 + a_{4i}^2 + a_{5i}^2 \end{bmatrix}_i = \begin{bmatrix} 1 & \rho_{1i}\theta_i & \rho_{2i}\sigma_i \\ \rho_{1i}\theta_i & \theta_i^2 & \rho_{3i}\theta_i\sigma_i \\ \rho_{2i}\sigma_i & \rho_{3i}\theta_i\sigma_i & \sigma_i^2 \end{bmatrix}.$$

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. Standard results regarding the log-normal distribution[11] imply the following expressions for the expectation and standard deviation of hourly earnings $\left(w_i^* = \exp(y_i^*)\right)$ for person *i*:

(5a) $$E(w_i^*) = \exp\left(\delta x_{yi} + \sigma_i^2/2\right),$$

(5b) $$S.D.(w_i^*) = \exp\left(\delta x_{yi}\right)\sqrt{\exp\left(2\sigma_i^2\right) - \exp\left(\sigma_i^2\right)}.$$

[11] Greene (1990), p. 64.

This means that in the presence of conditional heteroskedastity in logged earnings (i.e. α_2, α_4, α_5, $\neq 0$), a homoskedastic model is incapable of predicting not only the second, but also the first moment of the earnings distribution, underestimating the expected earnings for persons subject to above average wage variability. It also means that tests on δ do not suffice to test hypotheses regarding average actual (not logged) wages in a heteroskedastic world.

Next, we delineate the content of the main equations and the Cholesky matrix. Two 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 outcomes for different types of people, resulting in erroneous acceptance of the null of no diploma effects?

Logged wages, y_i^*, hours h_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 and hours, but not employment, as there are almost no unemployed union members or teachers in our sample. On this point, it is worth recalling that full employment amongst teachers is a prediction of the JSDH. Due to the small number of technical high school graduates with college education and no jobs, it was impossible to estimate separate employment effects of college for technical and high school graduates, and they are therefore assumed to have the same coefficients on college education.[12] Additionally hours and employment propensity are shifted by marital status and the interaction of marital status and gender. Hence we estimate the following conditional expectations functions:

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

[12] 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.

$$\delta x_{yi} = \delta^0 + \delta^M D_{Male} + \delta^R D_{Rural} + \delta^U D_{Union} + \delta^{Teach} D_{Teach}$$

(6b)
$$+\delta^{North} D_{North} + \delta^{South} D_{South} + \delta^E E + \delta^{E2} E^2$$
$$+ \sum_{l=p,j,h} \left(\delta^{sl} s_l + \delta^{Dl} D_l\right) + D_h\left(\delta^{sc}_{PF} s_c + \delta^{Dc}_{PF} D_c\right) + D_T\left(\delta^{DT} + \delta^{sc}_{PT} s_c + \delta^{Dc}_{PT} D_c\right).$$

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

It is clear from (4b) that the variables conditioning a_1 and a_2 will most strongly effect θ and σ respectively. Similarly, ρ_1, ρ_2 and ρ_3 can be conditioned through a_3, a_4 and a_5 respectively. There are 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 three characteristics. Similarly ρ_1, ρ_2, ρ_3 and θ are conditioned on the number of years of schooling.[13] Unionization was permitted to effect θ, σ and ρ_3 for obvious reasons. Finally, in keeping with the discussion of section II, σ was conditioned on the same variables as y_i^*, through a_2, in order to capture diploma effects in second moments. Hence, we fill out, $a_j = \alpha_j^0 + \boldsymbol{\alpha}_{\mathbf{j}}'\mathbf{x}_{\mathbf{ji}}$ $(j = 1,...5)$ to specify the following equations:

(7a) $$a_{1i} = \alpha_1^0 + \alpha_1^M D_{Male} + \alpha_1^R D_{Rural} + \alpha_1^U D_{Union} + \alpha_1^{Teach} D_{Teach} + \alpha_1^S Schooling$$

(7b)
$$a_{2i} = \alpha_2^0 + \alpha_2^M D_{Male} + \alpha_2^R D_{Rural} + \alpha_2^U D_{Union} + \alpha_2^{Teach} D_{Teach} + \alpha_2^E E + \alpha_2^{E2} E^2 +$$
$$\sum_{l=p,j,h} \left(\alpha_2^{sl} s_l + \alpha_2^{Dl} D_l\right) + D_h\left(\alpha^{sc}_{2,PF} s_c + \alpha^{Dc}_{2,PF} D_c\right) + D_T\left(\alpha_2^{DT} + \alpha^{Sc}_{2,PT} s_c + \alpha^{Dc}_{2,PT} D_c\right)$$

(7c) $$a_{3i} = \alpha_3^0 + \alpha_3^M D_{Male} + \alpha_3^R D_{Rural} + \alpha_3^S Schooling$$

(7d) $$a_{4i} = \alpha_4^0 + \alpha_4^M D_{Male} + \alpha_4^R D_{Rural} + \alpha_4^S Schooling$$

(7e) $$a_{5i} = \alpha_5^0 + \alpha_5^M D_{Male} + \alpha_5^R D_{Rural} + \alpha_5^U D_{Union} + \alpha_5^{Teach} D_{Teach} + \alpha_5^S Schooling\ .$$

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

Hypothesis tests:

This econometric structure allows us to test a variety of interesting hypotheses. We break these into four groups. First, this structure nests those of most Mincerian studies, thereby allowing us to test some of their assumptions. To the extent that these assumptions are rejected, estimated returns of many prior studies are inconsistent. We test the restrictions implied by a homoskedastic model ($\boldsymbol{\alpha}_{\mathbf{j}} = 0, \forall j$) by means of a likelihood ratio test (LRT) on the difference between the homoskedastic and heteroskedastic model. The resultant LRT statistic is distributed $\chi^2_{(33)}$. We also test for the presence of selectivity bias because y_i^* is undefined for the unemployed. The null hypothesis of no selectivity bias is imposed through the restriction $\alpha_3^0 = \boldsymbol{\alpha}_3 = \alpha_4^0 = \boldsymbol{\alpha}_4 = 0$, and the LRT statistic arising out of comparison with the full model is distributed $\chi^2_{(8)}$.

Second, we test the implications of the JSDH. 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 quantities $3\delta^{sh} + \delta^{Dh}$ and δ^{DT}. Given the presumed absence of downwards rigidity in wages accruing to either type of graduate, our job-specific diploma hypothesis suggests these quantities should be equal. To test the null more rigorously, we re-estimate the model with the equality imposed and test the restrictions via a $\chi^2_{(1)}$ LRT. In order to test the equality of actual (not logged) wages predicted, we impose the additional assumption that $3\alpha_2^{sh} + \alpha_2^{Dh} = \alpha_2^{DT}$, and compare the results to the full model by a $\chi^2_{(2)}$ 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 δ^{Teach} with $3\delta^{sh} + \delta^{Dh} + \delta^{sc}_{PF}$, (or $\delta^{DT} + \delta^{sc}_{PT}$). Given rigidities in the market for teachers, we expect the *normalista* to pay a wage premium under the JSDH. We formally test the restrictions $3\delta^{sh} + \delta^{Dh} + \delta^{sc}_{PF} = \delta^{DT} + \delta^{sc}_{PT} = \delta^{Teach}$ on the log wage premium using a $\chi^2_{(2)}$ LRT statistic relative to the full model. We test for a teaching premium in actual wages by adding the corresponding restrictions on α_2 and testing them jointly using a $\chi^2_{(4)}$ LRT statistic.

Third, we test for sheepskin effects. 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 $\delta^{Dl} > 0$ against the null $\delta^{Dl} \leq 0$. Similarly, the alternative hypotheses of positive sheepskin effects in second moments are tested via $\alpha_2^{Dl} > 0$. We also check the possibility that diplomas could improve access to jobs $\left(\beta^{Dl} > 0\right)$. Conversely, in the presence of explicit wage rigidities, the market for degreed employees might clear through more frequent unemployment $\left(\beta^{Dl} < 0\right)$. Tests for hours sheepskins are analogous.

Fourth, we examine the behavior of hours, seeking evidence of non-wage welfare effects of schooling. 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 the sub-sample obscures important results.

VI. Results:

We estimated the model delineated by equations (3), (4), (6) and (7) on the full sample of 16,675 workers.[14] The results are presented in table 2. 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 post-estimation 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, except that results are from estimates for the male and female sub-samples. Post estimation results for the other profiles we have looked at do not differ qualitatively. We also estimated the model on the following sub-samples: males, females, urban dwellers and young workers (those no older than 30).[15] The qualitative results are exactly the same in all samples, with the exception of those relating to the hours worked by men and women.[16]

---

[14] 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.

[15] The sample sizes for these cross sections were: 11,248, 5,427, 13,128 and 7,365 respectively.

[16] Results for sub-samples are available on request, as are post-estimation results for other profiles.

From the ‘hours’ column, we find that on average men work 2 hours more per week than women, married men work 4 hours more than bachelors, and married women work 4.6 hours less than unmarried women. We also find that men earn around 20% more than women on an hourly basis[17]. While this could be due to discrimination, we stress the importance of other, possibly complementary, interpretations. For example, as we will argue in detail below, Figure 3 *might* indicate that women choose different occupations from men, substituting flexibility in work hours for pay. However, with respect to Sheepskin effects and the returns to different schooling streams, the results were not gender specific.

The Cholesky terms a_1 and a_2 which condition the variances of hours and hourly earnings respectively are positive in every profile we examined. 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 one might expect, union membership reduces the variances of earnings and hours.

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 3 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 in this case lies in its ability to capture results in second moments, as well as to consistently test hypotheses regarding wages (not logged). 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. We conclude that Heteroskedasticity is statistically and economically relevant.

Similarly, the null that selection into employment does not bias results is rejected. Figure 4 demonstrates that for our reference profile ρ_2 is negative. This indicates that workers with greater 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 3, and reject it firmly. These results complement the findings of Alderman et. al. (1996) on this point in a rural Pakistani context. They find a positive correlation between the unexplained components of wages and the propensity to be engaged in wage

[17] For any binary log-wage intercept shifter, δ, the corresponding percentage increase in earnings is $e^{\delta} - 1$.

employment (as opposed to self-employment). Between their results and ours, it should be clear that the returns to schooling for employees cannot be presumed representative of the returns to any other demographic.

Our second set of tests verifies the predictions of the JSDH. 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%[18], 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. This is corroborated in the third row of Table 3 where we report a p-value of 0.88 from a $\chi^2_{(1)}$ test on the null hypothesis that the returns are equal. However, as equation (5a) suggests, a test in the predicted log wage could obscure actual wage differences only visible in the variance of log wages. A $\chi^2_{(2)}$ LRT on the null in actual wages is therefore conducted (Table 3, row 4), but 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, from (5a), that technical high-school graduates earn slightly more than formal graduates, as is visible in figure 2. This (albeit statistically insignificant) difference in wages is not captured by a homoskedastic Mincerian model.

Figure 2 provides visual confirmation of the above results. Note also that although the difference between wages for technical and formal graduates was insignificant at the end of high-school, and their Mincerian rates of return to college years were roughly equal at around 15%, the heteroskedastic Mincerian specification captures a large wage difference between them by the end of college. A homoskedastic model would not have revealed this difference between the wages earned by college graduates with formal and technical high-school backgrounds. This is further evidence of the insufficiency of a homoskedastic Mincerian specification for characterizing expected wages (not logged). Even a model that employs heteroskedasticity adjusted White errors would not capture this distinction. It is the structure of heteroskedasticity that matters.

Turning next to those with *normalistas*, we estimate a Mincerian rate of return over four years of 138%. Formal and technical high-school graduates with a year of college earn statistically indistinguishable four-year returns of 62.08% and 62.13% respectively. For rigor, we tested the results more formally through LRTs. In row 5 of table 3 we reject powerfully the null hypothesis that those with *normalistas* and those with either high school

---

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

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 α_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 Mincerian returns to technical high-school degrees, which are tied to jobs whose wages are more competitive, are the same as those to formal high-school. These observations exactly bear out the theoretical predictions that job-specific diplomas can generate sheepskin effects if, but only if, the wages paid by these jobs are shielded from competition. Finally, as noted in section III, given controls on the number teachers qualified, the JSDH predicts a non-negative employment sheepksin for teachers. We have already noted that *normalista* receipt guarantees employment in our sample.

Our third set of hypothesis tests look for sheepskin effects. In the 'logged earnings' column of table 2, 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 primary year boosts wages by a meager 0.6%. The 'a2' column provides compelling evidence of a positive primary sheepskin effect in the second moments of wages. We also see a negative primary sheepskin effect in the employment equation, and a positive primary sheepskin in the hours equation. 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. College graduation does carry an hours sheepskin.

These results conjure 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, explicitly require one. Thus, workers without a primary diploma may be restricted to more competitive and less formal markets. In the presence of a wage premium for diplomas and in the absence of restrictions on the number of primary diplomas to be granted, equilibrium across the two markets can only be sustained if the with-diploma market experiences higher unemployment or underemployment – 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. Because employing a working in Mexico's formal sector carries significant fixed costs, formal sector employers of degreed workers are expected to hire less workers for longer hours. This is fully consistent with the positive hours sheepskin and negative unemployment sheepskin. The positive sheepskin in the variance of log-earnings suggests, reasonably, that the job opportunities available to primary diploma holders offer more varied remuneration.

At this point the missing piece of the puzzle is the reason that the formal sector wage premium is not bid away. Certainly, if primary graduation required arduous or costly efforts a separating screening equilibrium could be a plausible answer. However, 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. Non-diploma holders are twice as likely to earn less than the minimum wage of five pesos per hour.[19] This rigidity could certainly sustain an equilibrium with unemployment and a modest wage premium in the degreed sector, as predicted by the JSDH

The finding that no other diploma effects exist is especially interesting 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.

It is worth mentioning that clustering is also fully consistent with the JSDH. Advertisements for job vacancies in Mexico frequently specify diploma requirements. This alone is enough to generate clustering. We have no reason to presume that the wages paid by said jobs are protected from falling due to increases in the supply of degreed workers. Therefore, any wage premium could be eroded by increases in the supply of degreed workers, resulting in clustering, but no diploma effect in the wage equation.

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 unmarried women 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

[19] Of those with no education beyond primary, 889 out of 3,367 non-diploma holders earn below the minimum wage, compared to 445 out of 3,029 diploma holders.

coincides well with results from Mehta and Villarreal (2004), which documents, using data from 2000, 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 offer regulated, full time work hours (40 per week). 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 ρ_3 (see figure 4) may suggest that there are basic necessities which workers with lower earnings potential must labor longer hours to afford, as suggested by Hernández-Licona (1997). These results suggest that the wage returns to education underestimate the private welfare benefits of schooling. More educated workers are able to enjoy more leisure in addition to higher consumption.

VII. Conclusions:

This paper proposes a simple institutional explanation of the value of diplomas in Mexico – that they are driven by wage rigidities and diplomas tied to specific jobs – and a methodology to test some implications of this hypothesis. We ask whether three other theories of diploma effects – screening, self selection, and productivity that increases discontinuously with years of schooling – reconcile with the data as well as our institutional hypothesis. We test for the divergent implications of these four theories for the relationships between diploma receipt and wages, the variance of wages, employment and hours worked.

The empirical evidence is fully consistent with the job-specific diploma hypothesis: premiums on diplomas appear to result from job-specific diplomas if, but only if, the wages paid by these jobs are protected. This conclusion is drawn from two sets of results.

First, despite the job-specificity of technical high-school degrees, technical high-school diplomas confer *exactly* the same wage premium as formal high-school diplomas. Teaching degrees, which lead to a specific job with a downwards rigid wage, confer more than twice the return of either high-school degree. Because the number of teachers certified is limited, teachers also experience nearly zero unemployment. As we expect of unionized employees, they work shorter than average hours.

Second, we find positive primary school sheepskins in both moments of the wage distribution, a negative primary sheepskin on employment, and a positive hours sheepskin. There is circumstantial evidence that primary diplomas are required for formal sector jobs, remuneration from which is constrained by minimum wage laws. Further, fixed costs of hiring formal sector workers are likely to lead employers to hire less workers and work them harder. Thus, institutional features and the JSDH could explain why primary graduates work longer hours, are unemployed more often, and earn more varied and substantially higher wages than their colleagues who drop out.

Our results therefore confirm the prediction of the JSDH that where labor markets are segmented according to diplomas, equilibrium is restored through unemployment amongst diploma holders when their wages are downwards rigid (primary graduates), and through restrictions on the number of degrees granted otherwise (teachers).

Pure screening and self-selection are unlikely explanations of primary diploma effects in Mexico, given the lack of significant barriers to primary school graduation. Non-linear productivity gains are plausible at the primary level but do not explain the unemployment, hours and wage variance sheepskins. Similarly, the theory of self-selection requires no non-Walrasian assumptions and therefore does not, on its own, explain the employment sheepskin effect. Our institutional hypothesis therefore stands out as the only considered theory that is fully consistent with all of our empirical results.

Interestingly, we find no other evidence of diploma effects, despite allowing for diploma and recipient heterogeneity. There is, however, indisputable evidence that students have disproportionately terminated their schooling immediately after receiving diplomas. These results are also consistent with the JSDH, which requires that if diplomas are tied to jobs whose wages are flexible, clustering without wage sheepskin effects will be the equilibrium outcome.

Nevertheless, these results demonstrate that studies on the returns to schooling must account very carefully for the institutional connections between the school system and the labor market, if their results are to have economically meaningful interpretations.

References:

Alderman, H., Behrman, J.R., Ross, D.R., and Sabot, R. (1996) 'The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market', *Oxford Bulletin of Economics and Statistics*, Vol. 58, pp. 29-55.

Arrow, K. (1973) 'Higher Education as a Filter', *Journal of Political Economy*, Vol. 2, pp. 193-196.

Azariadis, C. and Drazen, A. (1990) 'Threshold Externalities in Economic Development', *Quarterly Journal of Economics*, pp. 501-26.

Becker, G. (1967). *Human Capital and the Personal Distribution of Income,* University of Michigan Press, Ann Arbor, Michigan.

Behrman, J. and Deolalikar, A. (1991). 'School Repetition, Dropouts, and the Rates of Return to Schooling: The Case of Indonesia', *Oxford Bulletin of Economics and Statistics*, Vol. 53, pp. 467-480.

Belman, D. and Heywood, J. (1991). 'Sheepskin Effects in the returns to education: An Examination of Women and Minorities', *Review of Economics and Statistics*, Vol. 73, pp. 720-724.

Card, D. (2001). 'Estimating The Return to Schooling: Progress on Some Persistent Econometric Problems', *Econometrica*, Vol 69, pp. 1127-1160.

Chiswick, B. (1973). 'Schooling, Screening and Income', in L. Solmon and P. Taubman (eds) *Does College Matter?* Academic Press, New York.

El Universal, (2004) 'Ventajas del bachillerato tecnologico', Mexico City.

Fortin, N. and Lemieux T. (1997) 'Institutional Changes and Rising Wage Inequality,' *Journal of Economic Perspectives*, Vol. 11, pp. 75-96.

Galor, O. and Zeira, J. (1993) 'Income Distribution and Macroeconomics', *Review of Economic Studies*, Vol. 60, pp. 35-52.

Goldberger, A. (1993). *A Course in Econometrics*, Harvard University Press, Cambridge, Massachusetts.

Gottschalk, P. and Joyce, M. (1998). 'Cross-National Differences in the Rise in Inequality: Market and Institutional Factors', *Review of Economics and Statistics*, Vol. 80, pp. 489-502.

Greene, W. (1990). *Econometric Analysis*, Macmillan Publishing Company, New York.

Harris, J. and Todaro, M. (1970). 'Migration, Unemployment and Development: A Two-Sector Analysis,' *American Economic Review*, Vol. 40, pp. 126-42.

Heckman, J. (1970). 'Shadow wages, market prices, and labor supply,' *Econometrica*, Vol. 42, pp. 679-94.

Hernández-Licona, G. (1997) 'Oferta laboral familiar y desempleo en México', *El Trimestre Económico*, Vol. 64, pp. 531-568.

Hungerford, T. and Solon, G. (1987). 'Sheepskin Effects in the Returns to Education', *Review of Economics and Statistics*, Vol. 69, pp. 175-177.

Jaeger, D. and Page, M. (1996). 'Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education', *Review of Economics and Statistics*, Vol. 78, pp. 733-740.

Ljungqvist, L. (1993). 'Economic Underdevelopment: The case of a missing market for human capital', *Journal of Development Economics*, Vol. 40, pp. 219-39.

Mehta, A. and Villarreal, H. (2003). 'Mincerian Returns to Education in Mexico: Sheepskins and Sources of Heterogeneity', University of Wisconsin – Madison, Department of Agricultural and Applied Economics, Working Paper.

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

Psacharopoulos, G. (1985). 'Returns to Education: A further International Update and Implications' *Journal of Human Resources*, Vol. 20, pp. 583-604.

Riley, J. (1979). 'Testing the Educational Screening Hypothesis', *Journal of Political Economy*, Vol. 87, S227-S252.

Schady, N. (2003). 'Convexity and Sheepskin Effects in the Human Capital Earnings Function: Recent Evidence for Filipino Men', *Oxford Bulletin of Economics and Statistics*, Vol. 65, pp. 171-196.

Smith, P. and Metzger, M. (1998). 'The Return to Education: Street Vendors in Mexico', *World Development*, Vol 26, pp. 289-296.

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

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

(A2) $$\Pr(z = 0) = \Pr(z^* \le 0) = \Pr(\beta x_z + e \le 0) = \Phi(-\beta x_z).$$

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

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 $\left(h^*, y^*\right) \sim N\left(\mu_1, \Sigma_{11}\right)$ and $z^*\big|_{h^*,y^*} \sim N\left(\mu_2^*, \Sigma_{22}^*\right)$, where:

(A4a) $$\mu_1 = \begin{bmatrix} \gamma x_h \\ \delta x_y \end{bmatrix},$$ (A4b) $$\Sigma_{11} = \begin{bmatrix} \theta^2 & \rho_3\theta\sigma \\ \rho_3\theta\sigma & \sigma^2 \end{bmatrix}$$

(A4c) $$\mu_2^* = \beta x_z + \frac{\rho_1 - \rho_2\rho_3}{\left(1-\rho_3^2\right)}\left(\frac{h^* - \gamma x_h}{\theta}\right) + \frac{\rho_2 - \rho_1\rho_3}{\left(1-\rho_3^2\right)}\left(\frac{y^* - \delta x_y}{\sigma}\right)$$ and

(A4d) $$\Sigma_{22}^* = 1 - \left\{A^2 + 2\rho_3 AC + C^2\right\} \Big/ \left(1-\rho_3^2\right)^2; \;\; A = \left(\rho_1 - \rho_2\rho_3\right); C = \left(\rho_2 - \rho_1\rho_3\right).$$

Thus,

(A5) $$\Pr\left(z^* > 0 \middle| y^*, h^*\right) = \Phi\left(\mu_2^* / \Sigma_{22}^*\right)$$ and

(A6) $g\left(y^*, h^*\right)$ is the bivariate normal pdf characterized by $\left(\mu_1, \Sigma_{11}\right)$.

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

Appendix B: Formalization of a screening argument for second moment sheepskins

For simplicity we focus on graduation from a single diploma level and ignore the question of admission to that level. Potential employees have two characteristics: productivity, *a*, and a measure of academic performance on their diploma, *p*. The joint pdf of these variables is $\kappa(a, p)$. The corresponding marginal distribution of *p* is $\eta(p)$, with cumulative distribution $\mathrm{N}(p)$. If a diploma only states whether a candidate has passed or not, $\eta(p)$ would degenerate to the Bernouli distribution. The conditional distribution of *a*, given *p*, is $\tau(a|p)$. Following Arrow (1973), we make the stark simplifying assumption that employers cannot observe *a*. There is a cutoff p_0. Those with $p \geq p_0$ receive a diploma, and their value of *p* is visible to employers. If a student fails to reach this cut-off, or, aware of his impending poor performance, does not attempt the diploma, his academic performance remain unknown to employers.

In a competitive labor market, employers pay their best estimate of *a*. Thus, non-diploma recipients all receive the wage $w_{nd} = E(a|p \leq p_0) = (1/\mathrm{N}(p_0))\int_0^{p_0}\int_0^{\infty} a\kappa(a, p)\mathrm{d}a\mathrm{d}p$.

It follows that the variance of their wages is zero. A diploma holder with grade *p*, on the other hand, receives the wage $w_d = E(a|p) = \psi(p) = \int_0^{\infty} a\,\tau(a|p)\mathrm{d}a$. The pdf of *p*, conditional on diploma receipt is $\lambda(p) = \left[1\Big/\int_{p_0}^{1}\eta(t)\mathrm{d}t\right]\eta(p) = \eta(p)/\Pr(p \geq p_0)$. We make a *general screening assumption* that $\psi(p)$ is a strictly monotonic, once-differentiable function. Under this assumption we can derive, in the usual fashion, the pdf of w_g: $\omega(w_g) = \eta(\psi^{-1}(w_g))[\partial\psi^{-1}(w_g)/\partial w_g]/\Pr(p \geq p_0)$ from $\lambda(p)$. The expected wage for diploma holders is $E(w_g) = \int_{\psi(p_0)}^{\psi(1)} w_g \varpi(w_g)\mathrm{d}w_g$.

To ensure a positive first moment diploma effect for the average graduate, $E(w_g) = E(w|p \geq p_0) = E(a|p \geq p_0)$ $> w_{nd} = E(a|p \leq p_0)$, we must impose Arrow's *positive screening assumption.* This strengthens our *general* screening assumption, requiring that $E(a \mid p) = \psi(p)$ is upward sloping. Under the general screening assumption, the variance of diploma holders' wages is $V(w_g) = \int_0^{\infty}(w_g - E(w_g))^2\omega(w_g)\mathrm{d}w_g\Big/\Pr(p \geq p_0)$, which must be strictly positive, exceeding the variance of non-diploma wages (zero). Thus, a second moment diploma effect is implied by the conditions necessary to derive a first order effect.

Notice that if, as in Arrow's model, diplomas only carry binary information (pass or fail), then $\omega(w_g)$ is degenerate and no second order sheepskin should be observable. [20]

Table 1: Testable Implications of Theories of Diploma Effects

	Predicted Signs of Sheepskin Effects			
Theory of Diploma Value	**Wage**	**Wage-Variance**	**Employment**	**Hours**
Signaling	(+)	(+)	(?)	(?)
Non-linear Returns	(?) in general. (+) if academic effort is bunched in diploma years	(?)	(0)	(?)
Self-Selection	(+)	(?)	(0)	(?) in general. (+) if disutility from academic and professional work are positively correlated.
Job-Specific Diploma Hypothesis	(+)	(?) in general. Likely to be (+) if degreed workers qualify for non-degreed jobs.	(-) if diplomas are not restricted. (+) if diplomas are restricted	(?) in general. (+) if fixed costs are incurred to hire for jobs requiring degrees (-) if jobs requiring the degree are unionized

[20] It is certainly true that this entire argument ignores the type of market segmentation by which a student could, on completing a degree, become barred from certain jobs she was initially qualified for. However, this possibility, also assumed away by Arrow, seems more a feature of higher level degrees. We only find wage variance sheepskins at the primary level. We are grateful to Dr. Diana Weinhold for drawing our attention to this potential problem.

*Table 2. 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.
Intercept Shifters	Constant	**-0.2681**	0.0893	**48.3192**	0.7389	**1.5265**	0.0418	**0.6464**	0.0287
	Union			**-3.3826**	0.3022	**0.3585**	0.0144	**-0.1359**	0.0083
	Rural	**0.1248**	0.0421	**-1.2866**	0.2948	**-0.4221**	0.0200	-0.0220	0.0307
	Male	**-0.3223**	0.0426	**2.0601**	0.3596	**0.1837**	0.0141	**0.0684**	0.0187
	North	**-0.1515**	0.0368	-0.4308	0.2381	**0.1183**	0.0129		
	South	**0.1030**	0.0438	**2.1859**	0.2448	**-0.2320**	0.0141		
	Married	**0.8121**	0.0876	**-4.5949**	0.3592				
	Married Male	0.1232	0.0950	**8.5476**	0.4404				
	Experience	**0.0217**	0.0047	**0.1168**	0.0307	**0.0383**	0.0018	**-0.0118**	0.0009
	Experience squared	-0.0001	0.0001	**-0.0033**	0.0005	**-0.0005**	0.0000	**0.0002**	0.0000
Schooling Effects	Primary years	**0.3044**	0.0138	**-0.6235**	0.1299	0.0099	0.0074	**-0.0231**	0.0047
	Primary diploma	**-0.6380**	0.0853	**1.3634**	0.5632	**0.1234**	0.0316	**0.0802**	0.0196
	Junior-high years	0.0097	0.0510	0.2489	0.2806	**0.0635**	0.0159	0.0031	0.0081
	Junior-high diploma	0.1170	0.1425	-0.6409	0.7853	-0.0048	0.0448	-0.0190	0.0225
	Formal high-school years	0.0234	0.0850	-0.0276	0.3315	**0.1139**	0.0193	**0.0315**	0.0110
	Formal high-School diploma	-0.4036	0.2509	-0.3462	1.0184	-0.0203	0.0593	0.0306	0.0332
	3 year technical high-school degree	**-0.1550**	0.0754	-0.2695	0.4201	**0.3274**	0.0254	**0.1620**	0.0141
	4 year normalista			**-8.7169**	1.0477	**0.8741**	0.0558	**0.0874**	0.0317
	College years, post formal			**-1.1798**	0.1800	**0.1634**	0.0105	0.0099	0.0067
	College diploma, post formal			-0.4345	0.5014	**0.1594**	0.0263	-0.0419	0.0219
	College years, post technical			**1.5931**	0.7905	-0.0270	0.0468	-0.0562	0.0309
	College diploma, post technical			1.9451	2.6251	-0.1502	0.1407	0.1295	0.1149
	College years, all HS graduates	0.0257	0.0285						
	College diploma, all HS graduates	0.0317	0.1370						
		a1		a3		a4		a5	
		coeff.	std. err.	coeff.	std. err.	coeff.	std. err.	coeff.	std. err.
	Constant	**13.1337**	0.3284	**-10.8545**	0.6876	**-0.4724**	0.0345	**-0.5201**	0.0226
	Union	**-1.2506**	0.1577					-0.0220	0.0113
	Rural	**1.2793**	0.1701	0.1198	0.6393	**-0.2964**	0.0312	-0.0446	0.0268
	Male	**-0.7213**	0.1832	**5.1292**	0.5129	**0.0945**	0.0243	**0.1589**	0.0169
	Normalista	0.9133	0.6267					**-0.1801**	0.0406
	Schooling	**-0.0855**	0.0229	**0.7358**	0.0474	0.0034	0.0028	**0.0227**	0.0018

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

Table 3. Likelihood Ratio Test Results: Comparisons to the unrestricted model:**

	Null Hypothesis	Log Likelihood Value	LRT statistic	Degrees of Freedom	p-value
1.	Homoscedasticity	-80,159.91	1347.63	34	0.0000
2.	No self selection	-79,673.13	374.07	8	0.0000
3.	Formal - Technical log wage equality	-79,486.11	0.04	1	0.8371
4.	Formal - Technical wage equality	-79,487.28	2.38	2	0.3044
5.	Formal - Technical - Normalista log wage equality	-79,512.68	53.17	2	0.0000
6.	Formal - Technical - Normalista wage equality	-79,513.87	55.56	4	0.0000

**Log likelihood value = -79486.0913

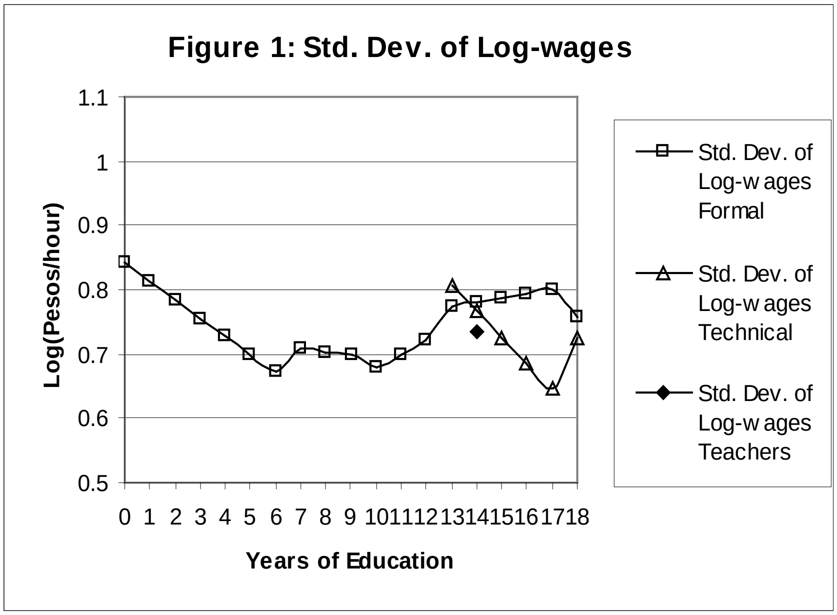
Figure 1: Std. Dev. of Log-wages
Log(Pesos/hour)
1.1
1
0.9
0.8
0.7
0.6
0.5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Years of Education
Std. Dev. of Log-w ages Formal
Std. Dev. of Log-w ages Technical
Std. Dev. of Log-w ages Teachers

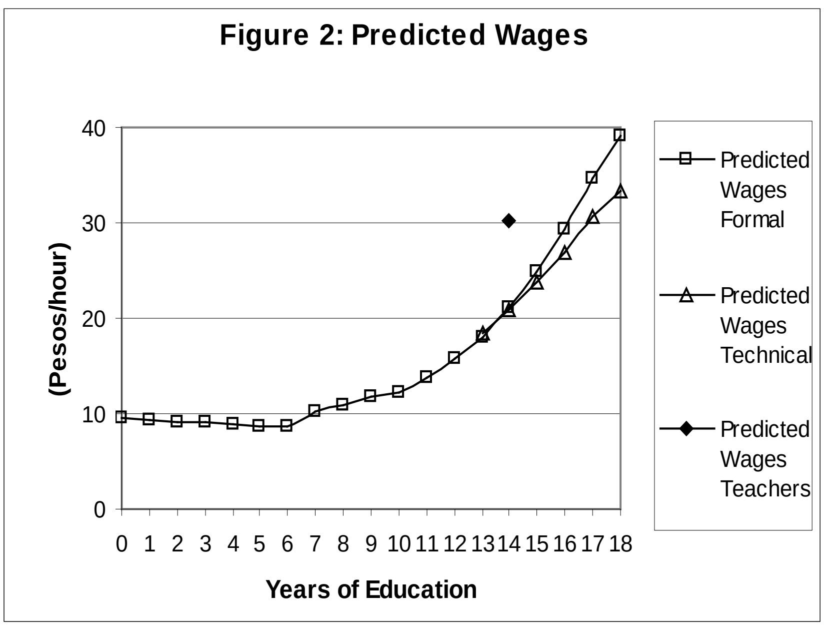
Figure 2: Predicted Wages
(Pesos/hour)
40
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Years of Education
Predicted Wages Formal
Predicted Wages Technical
Predicted Wages Teachers

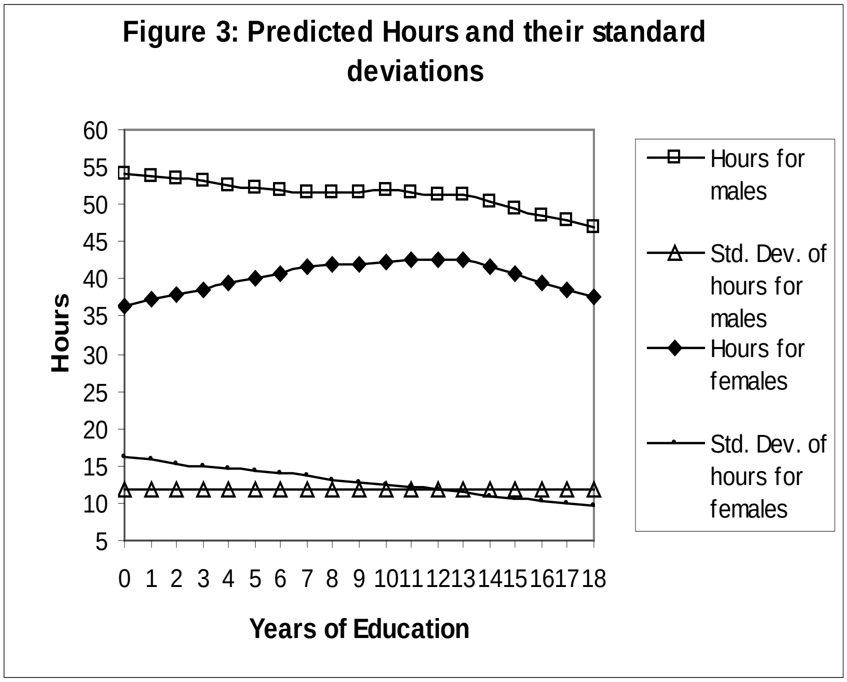
Figure 3: Predicted Hours and their standard deviations
60
55
50
45
40
35
30
25
20
15
10
5
Hours
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Years of Education
Hours for males
Std. Dev. of hours for males
Hours for females
Std. Dev. of hours for females

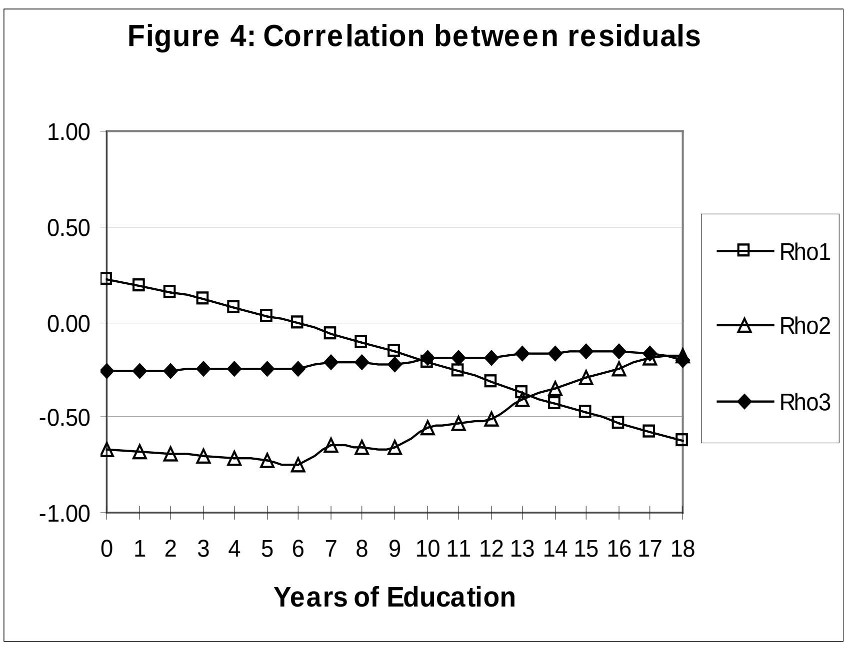
Figure 4: Correlation between residuals
1.00
0.50
0.00
-0.50
-1.00
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Years of Education
Rho1
Rho2
Rho3

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

Figure 5: Distribution of completed years of schooling.

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