

IZA DP No. 8736

Additive Kernel Estimates of Returns to Schooling

Deniz Ozabaci
Daniel J. Henderson

December 2014

Additive Kernel Estimates of Returns to Schooling

Deniz Ozabaci

University of Connecticut

Daniel J. Henderson

*University of Alabama
and IZA*

Discussion Paper No. 8736
December 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Additive Kernel Estimates of Returns to Schooling^{*}

In this paper, we employ a partially linear nonparametric additive regression estimator, with recent U.S. Current Population Survey data, to analyze returns to schooling. Similar to previous research, we find that blacks and Hispanics have higher rates of return on average. However, for married males, while non-Hispanic whites have lower returns on average, they typically possess the highest returns in the sample. For females, we are able to show that Hispanics have uniformly higher returns over non-Hispanic whites for the full sample. When we restrict our analysis to females whose highest level of education is a high-school diploma, we find average, but no longer uniformly higher returns. However, these uniformly higher returns resurface for college graduates.

JEL Classification: C14, J24

Keywords: additive, Mincer regressions, nonparametric, rate of return to education

Corresponding author:

Daniel J. Henderson
Department of Economics, Finance and Legal Studies
University of Alabama
Tuscaloosa, AL 35487-0224
USA
E-mail: djhender@cba.ua.edu

^{*} The authors would like to thank two anonymous referees, Subal Kumbhakar, Essie Maasoumi, Chris Parmeter, Sol Polachek and Le Wang for useful comments and suggestions as well as participants in presentations made at the State University of New York at Binghamton, Syracuse University, the University of Alabama and at the 2011 Midwest Econometric Group (University of Chicago). The R code used in this paper is available from the authors upon request.

1 Introduction

In the past four decades an immense amount of research has been devoted to the returns to schooling literature. This field is generally attributed to Mincer (1974) who proposed what is now commonly referred to as a Mincer regression model (or simply a Mincer regression). In this model, earnings are regressed on schooling and experience. The most common form in the literature includes log earnings on the left-hand-side and schooling, experience and the square of experience on the right-hand-side of the regression equation. A linear schooling variable suggests that an additional year of schooling gives the same percentage increase in earnings independent of the level of education. Sheepskin effects may be one of many reasons why this assumption may not hold true (Spence, 1973).

Many prominent authors have taken up this calling and proposed alternative models to capture possible nonlinearities and heterogeneity. Articles have proposed adding quadratic, cubic or quartic experience terms (e.g., Murphy and Welch, 1990). Considering the limitless possibilities that can be attempted, as expected, a consensus is not seen. This has led some to take an extreme stance and employ fully nonparametric regression (e.g., Heckman et al., 2008). The former estimators are believed to be inconsistent as they do not fully capture the underlying technology and the latter are deemed problematic because they do not place much structure on the function and result in too much variability.

In this paper we consider an intermediate approach. We want to relax the functional form assumptions on both schooling and experience, but we do not want a full blown nonparametric model. It is well known that nonparametric models suffer from the curse of dimensionality and a more parsimonious estimator may better explain the data.

The model we have in mind is a nonparametric additive regression model. These models allow regressors to enter nonparametrically, but separate from each other, hence avoiding the curse of dimensionality problem. However, in this strand of the econometric literature, little research has focused on obtaining the gradient of the conditional mean. In our particular problem, as well as many other problems in economics, we are interested in the partial effects.

We use the nonparametric additively separable model with interaction terms and additional linear controls to look at changes in earnings with respect to increases in schooling. This is our take on the classic Mincer (1974) model. This has been studied via kernel methods before (e.g., Heckman et al., 2006), but these methods are typically fully nonparametric and have a large amount of variability. We propose a more parsimonious model which leads not only to estimates with less variability, but also less extreme returns. In fact, we do not

find evidence of negative returns to schooling as was found in Henderson et al. (2011).¹

Our results confirm past research that heterogeneity in returns to schooling estimates exists both across and within standard sub-groups (race, sex and marital status). We also confirm past evidence that minorities receive higher rates of return to schooling on average than non-Hispanic whites.² However, by exploiting the heterogeneity, we are able to look deeper and find new results. First, for married males, while whites have lower returns on average, they typically possess the highest returns in the sample. Second, while minorities have higher mean returns, this may simply be a function of lower average levels of schooling. If we have a concave function (which is often assumed true for men, but not women, see Polachek, 1975), lower levels of schooling will result in higher returns, all else constant. We therefore further analyze our results by solely looking at individuals with fixed levels of education. Specifically, we look at individuals in each group whose highest level of education is a high school diploma and individuals whose highest level of education is a college degree. For example, whereas in the full sample of females (married or single), Hispanics uniformly possess higher returns to schooling (confirmed by first-order dominance tests), when we look at high school graduates, these uniform gains go away, but persist for college graduates. We further break down the results for Hispanic females based on ethnicity.

The remainder of the paper proceeds as follows: Section 2 gives a brief summary of the returns to schooling literature focusing on nonlinearities and parameter heterogeneity. Section 3 summarizes the conditional mean estimator of Kim et al. (1999), gives the gradient estimator and considers semiparametric extensions to the model. Section 4 describes the data, while Section 5 gives the empirical results. Section 6 concludes.

2 Nonlinearities and Heterogeneity in the Returns To Schooling Literature

It is impossible for us to summarize all of the developments in the returns to schooling literature in this paper. Those who are interested in this should consult the excellent survey by Polachek (2008). Instead, we are going to focus on past research analyzing functional form specifications and parameter heterogeneity within the Mincer framework.

Mincer (1974) himself proposed models that included higher-order terms for both school-

¹ For completeness, we also obtained empirical results with an additive local-linear estimator which are available by request. Most of the results are similar, but we do find some differences which we report in Section 5.2.3.

² Non-Hispanic whites will be referred to as whites hereafter.

ing and experience. Other earlier work on this topic includes Heckman and Polachek (1974) who look at several alternative models and conclude that the log earnings specification is appropriate. Polachek (1975) shows a nonlinear relationship for both men and women, but that the age-earnings path is non-monotonic for women. Later research by Card (1999), Heckman et al. (2006), Murphy and Welch (1990) and others argue that the quadratic specification of the experience term does not fit the data well and fails to capture the actual curvature of the age-earnings profiles. Card (1999) mentions that even a cubic specification fails to capture the data. Murphy and Welch (1990) conclude that a quartic specification for the experience term is better. Although this amount is relatively smaller than the cubic and quadratic alternatives, in their paper they still find some amount of bias. Belman and Heywood (1991), Park (1994) and others also criticize the linearity of schooling. Many of these studies start from the ‘sheepskin effect’ assumption and analyze the years that credentials are received and find nonlinearities around those years.

The complete separability assumption of the Mincer setting is also questioned (Card, 1999). It has been argued that the assumption of complete separability between experience and schooling is an important misspecification point (Card, 1999; Heckman et al. 2003). Basu and Ullah (1992) also question the specification assumptions made in the aforementioned papers. They attempt to ameliorate these problems by flexible nonparametric methods and argue that relaxing those restrictive assumptions provides more accurate age-earning relationships.

Beyond functional form, in his survey, Card (1999) considers heterogeneity as potentially important in determining returns to schooling. Griliches (1977) also expects the slopes of the earning functions to vary with different groupings. Other attempts analyzing heterogeneity can be found in Welch (1973) and Card and Krueger (1992a). Some of the possible sources of variation considered are school quality, family background and ability. Two recent examples with regards to the heterogeneous returns to schooling literature includes Koop and Tobias (2004) who analyze the heterogeneity issue using a flexible Bayesian approach. Henderson et al. (2011), using nonparametric kernel methods, focus on the same issue, but different from Koop and Tobias (2004), analyze within subgroup heterogeneity in addition to the heterogeneity between groups. They find significant heterogeneity both across and within groups.

3 Additive Nonparametric Estimation

Nonparametric methods are praised because they do not require *a priori* functional form specification and allow for interactions in an unknown way. However, one of the main drawbacks is the dreaded ‘curse of dimensionality’ problem (Stone, 1980). In Henderson et al. (2011), their returns to schooling estimates have substantial variability and while we agree that heterogeneity exists, it may be the case that this is exacerbated by the estimation method.³ In other words, their approach may make the population seem too heterogeneous. In fact, their research shows a substantial amount of negative returns and we wish to investigate this further.

By imposing an additivity constraint on the nonparametric regression model, we can reach an additively separable model which has the form

$$m(X) = c + m_1(X_1) + m_2(X_2) + \dots + m_d(X_d),$$

where the left-hand-side is the conditional mean function, and $m_s(X_s)$ for $s = \{1, \dots, d\}$ on the right-hand-side is a smooth function of the corresponding element X_s . Stone (1985) proves that the optimal rate of convergence in the additive model is independent of the number of regressors. This circumvents the curse of dimensionality problem. The two most common kernel-based methods used for estimating models of this kind are marginal integration (Tjøstheim and Auestad 1994; Newey 1994; Linton and Nielsen 1995) and back-fitting methods (Buja et al. 1989). The method we consider in this paper, the oracle estimator of Kim et al. (1999), is a combination of these two methods.

Related work in this literature includes Martins-Filho and Yang (2007) who examine the local-polynomial conditional mean estimator in the identically distributed case (noting that obtaining the gradient is straightforward); Linton and Mammen (2008) who consider a single-step conditional mean estimator in the weakly dependent case; and Cai (2002) who uses a two-step conditional mean estimator in the weakly dependent case that is similar to Linton (1997, 2000).

3.1 Conditional Mean

In the additive model, the conditional mean is modeled as a sum of smooth additive functions. Y is our left-hand-side variable and $X = (X_1, X_2, \dots, X_d)$ is a $d \times 1$ vector of regressors. For

³ These authors also use a cross-validation procedure which could possibly lead to a poor stochastic bandwidth choice.

ease of exposition, we consider the case where $d = 2$.⁴ With this restriction, our model becomes

$$Y = c + m_1(X_1) + m_2(X_2) + U,$$

where U is our mean zero additive error. For identification, it is typically assumed that $E[m_1(X_1)] = E[m_2(X_2)] = 0$ and hence $E(Y) = c$ is our intercept term.

If we were to directly nonparametrically regress Y on X_1 , we would have that

$$E(Y|X_1 = x_1) = c + m_1(X_1) + E[m_2(X_2)|X_1 = x_1].$$

The final term $E[m_2(X_2)|X_1 = x_1]$ creates a bias in estimation. In order to eliminate this bias, Kim et al. (1999) propose an instrument function, $w(x_1, x_2)$, such that

$$\begin{aligned} E[w(X_1, X_2)|X_1 = x_1] &= 1 \\ E[w(X_1, X_2)m_2(X_2)|X_1 = x_1] &= 0. \end{aligned}$$

If we have such an instrument, then it can be shown that

$$E[w(X_1, X_2)Y|X_1 = x_1] = c + m_1(x_1) \tag{3.1}$$

and this weighted conditional mean solves the bias problem and hence additive separability is achieved.

The exact instrument they propose is

$$w(x_1, x_2) = \frac{f_1(x_1) f_2(x_2)}{f(x_1, x_2)},$$

where $f_1(x_1)$, $f_2(x_2)$ and $f(x_1, x_2)$ are the marginal and joint probability densities, respectively. In practice, these can be estimated via kernel methods. We know that this is a valid instrument because

$$\int w(x_1, x_2) \frac{f(x_1, x_2)}{f_1(x_1)} dx_2 = \int f_2(x_2) dx_2 = 1$$

and

$$\int w(x_1, x_2) m_2(x_2) \frac{f(x_1, x_2)}{f_1(x_1)} dx_2 = \int m_2(x_2) f_2(x_2) dx_2 = 0.$$

To estimate (3.1), we can replace the densities defined in the instrument function by their

⁴ See Henderson and Parmeter (2014) or Li and Racine (2007) for the general case.

estimates. Formally, if we have n observations on Y , X_1 and X_2 , $\{Y_i, X_{1i}, X_{2i}\}_{i=1}^n$, our estimator of (3.1) is

$$\widehat{\gamma}_1(x_1) = \frac{1}{nh_0} \sum_{j=1}^n k\left(\frac{X_{1j} - x_1}{h_0}\right) \frac{\widehat{f}_2(X_{2j})}{\widehat{f}(X_{1j}, X_{2j})} Y_j, \quad (3.2)$$

where $k(\cdot)$ is a univariate kernel function with scalar bandwidth h_0 and our kernel estimators of the marginal and joint densities are

$$\widehat{f}_2(x_2) = \frac{1}{nh_0} \sum_{j=1}^n k\left(\frac{X_{2j} - x_2}{h_0}\right)$$

and

$$\widehat{f}(x_1, x_2) = \frac{1}{nh_0^2} \sum_{j=1}^n k\left(\frac{X_{1j} - x_1}{h_0}\right) k\left(\frac{X_{2j} - x_2}{h_0}\right).$$

We can think of (3.2) as a one-dimensional local-constant regression of a weighted Y on X_1 . Note that $\widehat{\gamma}_2(x_2)$ is similarly estimated by interchanging terms as necessary.

This estimator is less efficient than marginal integration and to obtain an efficient estimator we require an additional step, backfitting. The idea of the second-step originates from the oracle efficiency principle. Assume that we know $m_2(x_2)$. In order to estimate $m_1(x_1)$ efficiently, we first construct Y_1^{oracle} , which is a partial residual obtained assuming $m_2(x_2)$ is known:

$$Y_{1i}^{oracle} = Y_i - m_2(X_{2i}) - c.$$

If we knew Y_1^{oracle} , we could run a local-constant regression of Y_1^{oracle} on X_1 as

$$\widehat{m}_1^{oracle}(x_1) = \frac{\sum_{j=1}^n k\left(\frac{X_{1j} - x_1}{h}\right) Y_{1j}^{oracle}}{\sum_{j=1}^n k\left(\frac{X_{1j} - x_1}{h}\right)} \quad (3.3)$$

where h is the second-stage bandwidth (different from h_0) used for estimation of the smooth function of X_1 .

However, $m_2(x_2)$ and c are unknown, and thus this estimator is infeasible. We can replace c with the sample mean of Y (\bar{Y}) and $m_2(x_2) + c$ with $\widehat{\gamma}_2(x_2)$. The feasible version of Y_1^{oracle} (Y_1^{2-step}) is constructed as

$$Y_{1i}^{2-step} = Y_i - \widehat{\gamma}_2(X_{2i}). \quad (3.4)$$

This leads to our second-step estimator of $m_1(x_1)$, which is obtained by replacing Y_1^{oracle}

with Y_{1i}^{2-step} in equation (3.3).

Kim et al. (1999) show that under this approach, the local-polynomial version of the second-step estimator $\widehat{m}_1(x_1)$ has the same asymptotic distribution as the local-polynomial version of the oracle estimator. In other words, it is oracle efficient. Our focus here will be a local-constant estimator of the conditional mean and gradients. This choice is based purely on the plausibility of our resultant estimates in the empirical section.

The estimate of $\widehat{m}_2(x_2)$ is similarly obtained. With both estimates, we can obtain the fitted values for our additive regression as

$$\widehat{m}(X_{1i}, X_{2i}) = \bar{Y} + \widehat{m}_1(X_{1i}) + \widehat{m}_2(X_{2i}).$$

3.2 Gradients

In this section we derive the analytic gradients for the second-step estimator. Formally,

$$\frac{\partial \widehat{m}(x)}{\partial x_1} = \frac{\partial \widehat{m}_1(x_1)}{\partial x_1} \equiv \widehat{\beta}_1(x_1).$$

In order to estimate $\widehat{\beta}_1(x_1)$, we re-define our second-step estimator as

$$\widehat{m}_1(X_{1j}) = \frac{\frac{1}{nh} \sum_{j=1}^n k\left(\frac{X_{1j}-x_1}{h}\right) Y_{1j}^{2-step}}{\frac{1}{nh} \sum_{j=1}^n k\left(\frac{X_{1j}-x_1}{h}\right)} \equiv \frac{\widehat{g}_1(X_{1j})}{\widehat{f}_1(X_{1j})},$$

where $\widehat{g}_1(X_{1j}) \equiv \frac{1}{nh} \sum_{j=1}^n k\left(\frac{X_{1j}-x_1}{h}\right) Y_{1j}^{2-step}$ and $\widehat{f}_1(X_{1j}) \equiv \frac{1}{nh} \sum_{j=1}^n k\left(\frac{X_{1j}-x_1}{h}\right)$. With this notation, we can show that

$$\widehat{\beta}_1(x_1) = \frac{1}{\widehat{f}_1(x_1)} \left[\widehat{g}_1^{(1)}(x_1) - \widehat{m}_1(x_1) \widehat{f}_1^{(1)}(x_1) \right], \quad (3.5)$$

where $\widehat{f}_1^{(1)}(x_1) \equiv \partial \widehat{f}_1(x_1) / \partial x_1$ and $\widehat{g}_1^{(1)}(x_1) \equiv \partial \widehat{g}_1(x_1) / \partial x_1$ where

$$\widehat{g}_1^{(1)}(X_{1j}) = -\frac{1}{nh^2} \sum_{j=1}^n \left[k^{(1)}\left(\frac{X_{1j}-x_1}{h}\right) Y_{1j}^{2-step} \right] \quad (3.6)$$

where $k^{(1)}(\cdot)$ is the first-derivative of the kernel function with respect to x_1 .⁵

We wish to emphasize that we do not give any asymptotic results here. However, unre-

⁵ The second step can alternatively be estimated using local-linear least-squares regression (see Martins-Filho and Yang, 2007).

ported simulations show that the conditional mean and gradient estimators work quite well in both the i.i.d. and weakly dependent cases. For those interested in rigorous proofs in additive models, we suggest consulting Cai (2002), Kim et al. (1999), Mammen et al. (1999), Mammen et al. (2012), Martins-Filho and Yang (2007), Severance-Lossin et al. (1999) and Yang et al. (2003).

3.3 Interaction Terms

Additive models are attractive because they avoid the curse of dimensionality, but one potential drawback is they assume separability. It has been argued (Card, 1999) in Mincer models that an interaction term is needed between schooling and experience. Here we consider the case where the interaction is a known parametric function. Formally, we consider the model

$$Y = c + m_1(x_1) + m_2(x_2) + \psi(x_1x_2, \theta) + u,$$

where $\psi(\cdot)$ is a known function with unknown parameter θ . Manzan and Zerom (2005) and Schick (1996) provide methods to model the term linearly. For the linear interaction term, we use the technique described by Manzan and Zerom (2005). In the case of two regressors, our model becomes

$$Y_i = c + m_1(X_{1i}) + m_2(X_{2i}) + \theta X_{1i}X_{2i} + U_i, \quad i = 1, 2, \dots, n. \quad (3.7)$$

We can estimate this model by exploiting the weighted conditional expectation we used before. We do so by taking the weighted conditional mean with respect to each element in x as

$$\begin{aligned} E[w(X_1, X_2) Y_i | X_1 = x_1] &= E\{w(X_1, X_2) [c + m_1(X_{1i}) + m_2(X_{2i}) + \theta X_{1i}X_{2i} + u_i] | X_1 = x_1\} \\ &= c + m_1(X_{1i}) + \theta E[w(X_1, X_2) X_{1i}X_{2i} | X_1 = x_1] \end{aligned} \quad (3.8)$$

and

$$\begin{aligned} E[w(X_1, X_2) Y_i | X_2 = x_2] &= E\{w(X_1, X_2) [c + m_1(X_{1i}) + m_2(X_{2i}) + \theta X_{1i}X_{2i} + u_i] | X_2 = x_2\} \\ &= c + m_2(X_{2i}) + \theta E[w(X_1, X_2) X_{1i}X_{2i} | X_2 = x_2]. \end{aligned} \quad (3.9)$$

Subtracting (3.8) and (3.9) from (3.7) gives us

$$\begin{aligned}
& Y_i - E[w(X_1, X_2) Y_i | X_1 = x_1] - E[w(X_1, X_2) Y_i | X_2 = x_2] \\
= & -c + \theta \left\{ \begin{array}{l} X_{1i} X_{2i} - E[w(X_1, X_2) X_{1i} X_{2i} | X_1 = x_1] \\ -E[w(X_1, X_2) X_{1i} X_{2i} | X_2 = x_2] \end{array} \right\} + u_i,
\end{aligned}$$

where the intercept is $(1 - d)c = -c$ when $d = 2$.

After estimating $E[w(X_1, X_2) Y_i | X_1 = x_1]$, $E[w(X_1, X_2) Y_i | X_2 = x_2]$, $E[w(X_1, X_2) X_{1i} X_{2i} | X_1 = x_1]$ and $E[w(X_1, X_2) X_{1i} X_{2i} | X_2 = x_2]$ each via one-step marginal integration, the parameter θ can be estimated via least-squares. Once an estimate of θ is obtained, we can move the estimated parametric term to the left-hand-side and the additive functions are again estimated via the additive estimator. For further details see Li and Racine (2007) or Manzan and Zerom (2005).

3.4 Partially Linear Model

In addition to adding interaction terms, we may also be interested in including additional explanatory variables linearly. For example, in Mincer models, it is common to include control variables such as region of residence or number of children. These variables can be used to help control for sub-group heterogeneity. Once we add these control variables, our model (with the interaction term) can be written as

$$Y_i = c + m_1(X_{1i}) + m_2(X_{2i}) + \theta X_{1i} X_{2i} + Z_i \gamma + U_i, \quad i = 1, 2, \dots, n, \quad (3.10)$$

where Z_i is a $1 \times q$ vector of control variables with $q \times 1$ parameter vector γ . All other variables and parameters are the same as before.

This model is estimated using a three-step approach. First, we obtain conditional expectations via kernel regression for our estimate of γ , a la Robinson (1988). Second, we estimate the interaction term $\hat{\theta}$ as above (Manzan and Zerom, 2005) where our left-hand-side variable is now $Y_i - Z_i \hat{\gamma}$, where $\hat{\gamma}$ is our estimate of γ from the first-step. Third, now that we have estimated both θ and γ , we can use our additive estimator for a regression of $(Y - \hat{\theta} X_1 X_2 - Z \hat{\gamma})$ on X_1 and X_2 .

4 Data

In this paper we use the March 2010 U.S. Current Population Survey (CPS) data which was obtained through the IPUMS-CPS database (Integrated Public Use Microdata Series). The CPS is a monthly U.S. household survey with a size of over 50,000 households and it is under control of the U.S. Census Bureau and the Bureau of Labor Statistics. IPUMS-CPS is the organized version of March-CPS, by the Minnesota Population Center (King and Tertilt, 2003; King et al., 2010).

The primary variables of interest are years of schooling, potential experience and earnings. There are a couple of important points when it comes to data selection in the Mincer framework. First, how do we define the earnings variable's duration and form? The popular form adopted in the literature is logarithmic earnings on the left-hand-side (Mincer, 1974). Heckman and Polachek (1974) discuss several functional forms and conclude that the so-called semi-logarithmic function performs best. Card (1999) supports this popular form considering the normal distribution proximity of the left-hand-side variable and the so-far success of the transformed functions. Following those papers' lead, we also model our left-hand-side variable in the logarithmic form.

The other issue regarding earnings is the choice of time frame. Card (1999) states that since individuals with higher schooling will be prone to working more, they also will be earning more. Therefore, any returns to schooling estimation based on weekly or annual earnings will be higher than hourly counterparts. Following the annual earnings description by Card (1999) and imitating the CPS Stock and Watson (2007) data, we perform our estimations using average hourly earnings.

A second issue concerning data for the Mincer (1974) regression model is the reporting of the schooling variable. About three decades ago, the CPS database changed its reporting to a degree completion system and this made it difficult to estimate the usual Mincer type models (Card, 1999). The March-CPS (after 1990) has a mixed version. For those that have degrees under high school, they record the highest school grade completed, but only report intervals. After high school, only degrees/diplomas obtained are reported. In order to solve this discrete data problem, we follow the midpoint approach (Lemieux and Card, 2001).

Here we hope to examine two types of heterogeneity: across pre-specified sub-groups and within pre-specified sub-groups. We divide our sample into 12 distinct groups: white married male, white married female, white single male, white single female, black married male, black married female, black single male, black single female, Hispanic married male,

Hispanic married female, Hispanic single male, and Hispanic single female workers. We note here that we prefer to create subgroups with respect to gender, ethnicity and marital status. We consider gender to be able analyze female and male workers separately, considering the different labor market conditions they face. Ethnicity is included as one of the main source of heterogeneity and this is parallel to the examples in the literature. Marital status on the other hand is included, to be able to capture any existing marriage premium and to consider the different labor supply characteristics of single and married workers. The alternative would be to include dummy variables and/or discrete regressors. The former assumes the additive nonparametric functions are the same between groups and the latter would require theory which is beyond the scope of the paper.

The descriptive statistics for these groups, including sample size used for each, can be seen in Table 1. We run separate parametric and nonparametric regressions for each sub-group to allow for parameter heterogeneity and/or different functional forms across sub-groups.

The one point we wish to emphasize here is the difference in the average (or median) number of years of schooling. We see that the median years of education is generally lower for Hispanics. If we, for example, have a concave function, this would suggest higher returns to an additional year of education for Hispanics on average, all else constant. We will return to this point in Section 5.2.2.

We also include a number of control variables. In order to maintain a relatively large sample, we include a dummy variable indicating whether an individual is not a citizen, a dummy variable indicating whether the individual has a child under age 5, the number of siblings and controls for geographical region. Region is represented by three dummies for North East, Midwest and South. We also considered other controls such as mother's education. These resulted in no major qualitative differences in the returns to schooling results, but substantially reduced the sample size. The results from these additional regressions are available from the authors upon request. These variables are included in our regressions in order to control for individual and demographic characteristics. They are chosen from those which are already employed in the existing literature. Citizenship is expected to influence the schooling effects as it directly effects employment decisions. Geographical location and number of young children on the other hand are expected to control for household characteristics, whereas the number of siblings is a control for family/background.

Finally, following the literature, we restricted our sample following Card (1999) (as his approach is the least restrictive version we found in the literature). Specifically, we only include full-time workers age 16-66 with positive hours of work, experience and education.

Further, those with average hourly earnings below \$2 and above \$150 are dropped. The data set used in the regressions is available from the authors upon request.

5 Returns to Schooling Estimates

In this section we focus on the returns to schooling estimates from both the parametric and nonparametric models. We first consider linear and nonlinear parametric models. These results are available in Table 2. We then formally test for correct parametric specification using the test introduced by Ullah (1985). After rejecting each parametric specification, we turn to our preferred specification. We give a summary of the returns to schooling from the nonparametric procedure in Table 3. This table can show heterogeneity across groups, but hides the heterogeneity within sub-groups. Given that we obtain an estimate of returns to schooling for each individual, we analyze the empirical cumulative distribution functions (ECDF) constructed with these estimates to look at returns both within and across sub-groups. We look at the ECDFs for all individuals in a specific sub-group as well as look at ECDFs constructed with just people of a given education level. Specifically, we look at the empirical distributions for people in each sub-group whose highest level of education is a high school diploma as well as those whose highest level of education is a college degree. We end this section with a note on endogeneity.

Before diving into the results, we need to discuss an issue with respect to female sub-groups. In the case of females, estimating earnings functions can be problematic when we use the standard metric to obtain the experience variable ($\text{Age} - \text{schooling} - 6$). This is a problem because there may be gaps in the labor force participation of women mainly due to child-rearing. One solution to this issue comes from Buchinsky (1998) and Martins (2001). They add interaction terms for experience and the total number of children to the main earnings function. We follow the same approach.⁶ Thus, for female subsamples, we have two additional interaction variables; total number of children times experience and total number of children times the square of experience. This helps control for gaps in potential experience, but does not fully solve the problem. Therefore, we should be careful when interpreting differences between men and women.

⁶ Note that the aforementioned papers are primarily concerned with sample selection problems.

5.1 Parametric

A variety of extensions to the classic Mincer (1974) specification has been proposed in the literature. Each one attempts to improve upon the former. Here we consider several parametric models, not only to compare to the previous literature, but also as an intermediate step between the basic parametric specification and nonparametric methods.

We consider six separate specifications from the literature (from an endless possibility of models) which include higher-order terms of schooling and experience, but are still linear in parameters so that we can use least-squares techniques. The list of models considered are

Model	Parametric Specification
1	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + Z_i \gamma + U_i$
2	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 E_i^3 + Z_i \gamma + U_i$
3	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 E_i^3 + \beta_5 E_i^4 + Z_i \gamma + U_i$
4	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 E_i^3 + \beta_5 S_i^2 + Z_i \gamma + U_i$
5	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 E_i^3 + \beta_5 S_i^2 + \beta_6 S_i^3 + Z_i \gamma + U_i$
6	$\ln(\text{earnings}_i) = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \beta_4 E_i^3 + \beta_5 S_i^2 + \beta_6 S_i^3 + \beta_7 S_i E_i + Z_i \gamma + U_i$

where $\ln(\text{earnings})$ is the logarithm of average hourly earnings, S is schooling, and E is potential experience. Z includes citizenship status, whether the person has children under age 5, number of siblings and regional dummies. In the case of females, it also includes the interaction terms discussed above. The parameter β_0 is the intercept, β_1 through β_7 are the coefficients associated with various incarnations of schooling and experience, and γ is a vector of coefficients associated with the control variables.

The first model is the standard Mincer (1974) specification which is the most popular in the applied literature. The second model adds an additional cubic experience term as discussed in Card (1999). Model 3 adds an additional quartic experience term, similar to Murphy and Welch (1990). We also estimate several non-standard models. The fourth model includes a quadratic schooling term and Model 5 adds a cubic schooling term. Each attempt to capture potential nonlinearity in the schooling variable. Our final parametric model nests each of the previous models as it includes each of these higher-order terms as well as an interaction term between schooling and experience.

The results from the parametric models can be found in Table 2. Here we report the median return to schooling across the population along with its associated standard error below the estimate in parentheses.⁷ In Models 1-3, schooling enters linearly and hence the

⁷ We choose to solely focus on the returns to schooling variable: $\partial \ln(\text{earnings}) / \partial \text{schooling}$. The other

estimate in the table is simply the coefficient β_1 .

For the standard specifications, the results are similar to what is found in the literature. If we look at the models where we add higher-order schooling terms (Models 6 and 7), we often find median results which are unsettling. There are several cases where the median return to schooling is negative (we are likely imposing too much or the wrong type of nonlinearity). Even though Henderson et al. (2011) found the presence of some negative returns, their median results were positive and significant. In addition, we find some returns to schooling which are above unity. We are unaware of any results in the literature which make such a finding nor would we expect one. In short, adding higher-order terms does not necessarily solve the nonlinearity concerns and sometimes leads to undesired results.

Even though some of these nonlinear parametric models clearly seem to be misspecified, we wish to perform formal tests for correct parametric specification. For each of the six models, for each of the twelve sub-groups, we perform the test developed by Ullah (1985). In each case, we test the parametric model versus our preferred nonparametric model. The p-value for each test, obtained via a wild bootstrap, was zero to four decimal places. In other words, we reject the null that each parametric model is correctly specified. Each test points toward our semiparametric specification.⁸ We now focus our attention on our preferred semiparametric specification.

5.2 Semiparametric

Here we present the results from our preferred semiparametric specification. Specifically, we consider the specification in equation (3.10). Our model becomes

$$\ln(\text{earnings}_i) = c + m_1(S_i) + m_2(E_i) + \theta S_i E_i + Z_i \gamma + U_i,$$

where c is the intercept term, S is schooling and shows up in the unknown function $m_1(\cdot)$ and linearly interacted with experience (E). Experience enters the equation through the unknown function $m_2(\cdot)$ and is interacted linearly with schooling, and for females, with the total number of children. Z again includes citizenship status, whether the person has children under age 5, number of siblings and regional dummies. We run separate semiparametric regressions for each of our twelve sub-groups. We first report the median return to schooling

estimates, such as for the control variables, are in line with the literature and are available upon request.

⁸ We also performed specification tests for our preferred semiparametric model versus a nonparametric alternative and failed to reject the null that the semiparametric model was correctly specified in 10 of the 12 cases - the exceptions being white married males and single black males.

for each sub-group in Table 3, similar to what we did in the parametric sub-section. We do this for the pooled sample as well as for given levels of schooling (i.e., high school and college graduates). However, here we also look at the ECDFs of the returns to schooling estimates. These are given in Figures 1-3. With these figures, we examine both heterogeneity within groups and across groups. Further, we use Kolmogorov-Smirnov tests, similar to those in Eren and Henderson (2008), to look for cases of stochastic dominance between the estimated returns to schooling across groups where applicable.

Before we discuss our results, we should spend some time to talk about bandwidth selection. Data driven bandwidth methods (e.g., least-squares cross-validation) can be problematic in this sense because they are stochastic sequences (Martins-Filho and Yang, 2007). In this study, we use the plug-in bandwidths proposed in Kim et al. (1999) which require the first-step to be under-smoothed.

5.2.1 Median Returns

The median returns to schooling for each sub-group are given in Table 3. There are several points worth mentioning here. First, the median returns for each group are substantially larger than what we saw in the standard parametric models (1-3). This result is not too surprising given that it is similar to what was found for black and white men in Henderson et al. (2011). Note that in their study, they only examined males and do not report results for Hispanics.

Second, black and Hispanic workers have higher median returns than whites. The literature so far has come to a consensus that individuals from seemingly disadvantaged groups tend to have larger returns to education (Card and Krueger, 1992b; Welch, 1973). As Card (2001) argues, the tight financial constraints of these groups and the likelihood of high education costs can be an explanation for these relatively higher values. Brand and Xie (2010) argue that the difference between the sub-population returns can be explained through advantaged group characteristics. People from more advantageous groups tend to have higher education levels. It is argued that this increases the likelihood of smaller returns for whites (Brand and Xie, 2010). Henderson et al. (2011) confirm this result, but they also add that a high level of heterogeneity is present even within the so-called advantageous groups. We will exploit this later.

Regarding the difference between median returns of black and Hispanic workers, we do not see a clear pattern. We will see that these median returns can be potentially deceiving when we look at the ECDFs.

Finally, we want to emphasize that the standard errors for our nonparametric estimates are relatively small. While these do not appear to be all that much different from those in the parametric models, they are much smaller than what we would expect in a fully nonparametric specification. If we take the model of Henderson et al. (2011) as our comparison, the standard errors on our median returns to schooling estimates are several times smaller than their (unreported) standard errors (and their model only includes three regressors – schooling, experience and a dummy for top-coded). If we were to estimate their model using our full set of regressors, we would expect even more variation in their estimates. That being said, we should mention that they use least-squares cross-validation to estimate their bandwidths and this may also help explain the additional variability in their model.

5.2.2 Empirical Distributions of Estimates

As we have extensively emphasized, heterogeneity is an important factor. Using mean or median results to formulate policy could have a detrimental impact on individuals in the tails of the distribution if heterogeneity is present. Here we go beyond mean and median estimates and look at the entire distribution of estimates for each sub-group. This will accomplish at least two things. First, we can see the heterogeneity within each sub-group. Second, we can see whether or not there exists uniform dominance between sub-groups. We will look at the latter formally with Kolmogorov-Smirnov tests. For the critical values of the test statistic, we use bootstrap procedure (Henderson and Maasoumi, 2014).

We consider two sets of ECDFs. The first set will look at all individuals within a particular sub-group. This will allow us to examine differences between groups based on race and marital status. However, as we noted in Section 4, some groups have higher average levels of education than others and this may affect the returns to schooling estimates. We therefore consider two other figures. The first will look at all individuals within each sub-group whose highest level of education is a high school diploma. The second will look at all individuals within each sub-group whose highest level of education is a college degree. This will allow us to detach the level of education from the return and hence will let us make ‘more comparable’ estimates between the sub-groups.

All Education Levels In Figure 1 we have four panels (a-d) with three ECDFs in each panel. Each panel corresponds to a separate sex-marital status combination and each of the ECDFs within the panel correspond to a particular race.

In panel (a) we have the ECDFs which are constructed from the returns to schooling

estimates for black, Hispanic and white married males. We want to emphasize at least two points in this panel. First, we see the heterogeneity in the estimates. We see some individuals who have very low returns to education (near zero) and some individuals which have very high returns. This is similar to the results of Henderson et al. (2011). The second point is more novel. While we see that the average return to schooling for blacks and Hispanics is larger than that of whites, most of the highest returns are gained by whites. Graphically, this is shown by the intersection of the ECDFs. These intersections suggest that there is no first-order dominance relation between whites and blacks or Hispanics and this is confirmed by the Kolmogorov-Smirnov tests which have p-values that are zero to four decimal places.

Although there are no first-order dominance relations with respect to whites, we do find evidence of first-order dominance of Hispanics over blacks (p-value = 0.7663). At the lower tail it is difficult to differentiate the two empirical distribution functions, but we see clear separation at the higher returns.

To further dissect the results, we examine the attributes of individuals both above and below where the empirical distributions cross. In the higher return group, we find slightly lower levels of schooling. Some evidence, albeit weak, to suggest that higher levels of education lead to lower returns. In addition, we find higher average levels of experience in the high return portion. We tried, but were unsuccessful at determining which occupations lead to the highest returns.

Panel (b) of Figure 1 is also interesting, but different from that of panel (a). We again find heterogeneity for each sub-group of married females, but now find two dominance relations. Here we do not see the same gains for whites at the higher end that we saw in panel (a). In fact, there is some graphical evidence to suggest that married Hispanic women have the highest returns. Instead of relying on the figures, which are often difficult to decipher, we prefer to consult the Kolmogorov-Smirnov tests. We find that both blacks and Hispanics first-order dominate whites with respect to returns to schooling (p-values = 0.6312 and 0.9810, respectively).

The results for single men can be found in panel (c). Here we find similar returns in the left tail for all races, but more mass for white males in the right tail of the distribution. This is similar to what we found for married white males, but not to the same extent. We find no first-order dominance relations between any two groups.

Finally, for single females, the ECDFs for each race can be found in panel (d). Here we find strong evidence to suggest that the returns for Hispanic females are uniformly greater than those for black or white single females. The ECDF for the returns to schooling for

single Hispanic females appears to lie to the right of the other two curves. This result is confirmed by the Kolmogorov-Smirnov test where we have p-values of 0.9992 and 0.9901 and hence fail to reject the null of first-order dominance for larger Hispanic returns over both black and whites, respectively. Note that we do not find a dominance relation for blacks over whites or vice versa. It is obvious from the figure that these curves intersect.

Fixed Levels of Education The results in Figure 1 help us gain many insights, but they may also hide some features. As we mentioned in Section 4, Hispanics on average have lower levels of schooling in our sample. This could be one explanation for their higher average returns. In this section, we look at the returns to schooling for fixed levels of education of interest. Specifically, we take the results from above, but select individuals from each sub-group with a given level of schooling. We consider two groups: those whose highest level of schooling is a high-school diploma and those whose highest level of schooling is a college degree. The analogous figures to that in Figure 1 can be found in Figures 2 and 3 for high school and college, respectively.

Panel (a) of Figure 2 looks a lot like the corresponding panel in Figure 1. Here we have the returns to schooling for married males with a high school diploma. We again find smaller average returns to schooling for whites (0.207 versus 0.221 and 0.219 for blacks and Hispanics, respectively). We also find that the curves cross so that whites often obtain the highest returns to schooling. If we turn our attention to panel (a) of Figure 3, we find the same result. In short, this phenomenon appears to hold for both the full sample as well as the high school graduate and college graduate samples.

For the married female group, we find that the dominance relation no longer exists for Hispanics for high school education, but does for college graduates. For blacks, there are no first-order dominance relationships over whites for fixed years of schooling. In fact, we find that the average returns to schooling are actually relatively similar across the two schooling distinctions. This downplays the higher returns for ‘less advantageous’ groups, but does not eliminate it. For the high school graduate case, the average return for married females for blacks, Hispanics and whites are 0.219, 0.228 and 0.214, respectively. For the college degree case, those numbers are 0.165, 0.172 and 0.178. One benefit of pointing out these numbers is to show that higher levels of education are associated with lower returns.

The single male case also does not appear to differ among the schooling levels. In fact, panel (c) in each figure looks similar. We see that the empirical distributions are difficult to distinguish from one another at lower returns, but whites appear to have more gains in the

higher returns portion. In other words, the right tail of the empirical distribution for whites is thicker.

The most interesting case is for single females. In Figure 1, we find uniform dominance for Hispanic single female returns to schooling over other groups. When we focus solely on high school graduates, we see the ECDF for Hispanics is to the right of the others for lower levels of returns to schooling, but less than that of whites for higher returns to schooling. In fact, this result looks much more like that for married men. Although it is to a lesser degree, it is clear that there is no longer a uniform dominance for Hispanic single females. The p-value for the Kolmogorov-Smirnov test is relatively small (0.1912).⁹ For college graduates, we continue to see lower returns to be more prominent for Hispanics, but now see that the higher returns are similar for Hispanics and whites. That being said, we again find a dominance relation (p-value = 0.9411).

In summary, some of the results for the full sample continued to hold, but some did not. For example, married males looked very similar in each of the figures. On the other hand, the uniform dominance we found for Hispanic single females over the other two groups did not hold for high-school graduates, but continued to hold over whites for college graduates. Similarly, for Hispanic married females, we found uniformly higher returns for the full sample and the college graduate sample over that of white married females. The last two sets of results suggests that we should devote future research efforts towards finding the determinants of high returns to college degrees for Hispanic females.

5.2.3 Local-Linear Least-Squares

We have mentioned our preference to go with the local-constant estimator, but given the benefits of local-linear estimators relative to local-constant estimators it makes sense to examine the local-linear results as well. Appendix B gives the local-linear results for our application. In general, the main findings do not change much. We find higher returns for minorities males, but no cases of dominance. We also find higher returns for minority women and the same dominance relations as for the local-constant case.

As for the differences between the local-constant and local-linear estimates, we also found evidence for stochastic dominance of returns to schooling for black women with college degrees over white women with college degrees. For men, the primary difference is that we no longer find the larger returns for white married males for the local-linear case.

⁹ In the stochastic dominance literature it is common to use the cut-off value of one-half (0.50) for p-values (e.g., Eren and Henderson, 2008).

One difference that occurred over most subgroups was the tendency for negative returns from a portion of the sample. This result is similar to what was found in Henderson et al. (2011) who also used a local-linear estimator. However, Henderson et al. (2011) used a cross-validation routine which often leads to more variability in the estimates. That being said, given the relatively small proportion of negative estimates (for most cases), it is unclear whether this difference is economically meaningful.

In summary, we find that the majority of our findings do not change when using a local-linear estimator. This not only gives us some faith in the empirical exercise, but also for the performance of our proposed local-constant gradient oracle estimator.

5.2.4 Hispanic Females

Given the interesting results for Hispanic women, we decided to look a bit closer. We could think of many ways to parse the results, but focus on ethnicity. It has been argued (Landale et al., 2006) that excluding geographical origin in a study of Hispanics will fail to capture important cultural differences. Mexicans, Puerto Ricans and Cubans are the three major Hispanic groups with the largest US-born populations (Duncan et al., 2006). We will look at these three along with a generic category for Central and South America.

We first divide the estimated returns to schooling values among the individual's (ancestral) country of origin and we further divide these sub-samples via marital status and among different education levels as we did above. Table 4 gives median and mean education levels for each group: Mexicans, Puerto Ricans, Cubans and South/Central Americans. The results are in line with the findings from the literature. Cubans tend to have notably high levels of education and Mexicans tend to have the lowest schooling levels (Schneider et al., 2006). The descriptive statistics for Mexicans are similar to those of the pooled sample as they constitute over half the sample.

The returns to education for both married and single Hispanic females are given in Table 5. For each education level, Mexican and Cuban married females get smaller returns than the related single female subgroups. The opposite result holds true for Puerto Rican and Central/South American women. This result deserves more attention.

It is also interesting to consider the impact of citizenship (noting that Puerto Rico is a US territory). Noting that we have a control for this in our regressions, for example, we see that nearly 60% of married Mexican female college graduates are citizens while only 50% are so for high school graduates. Around one quarter of married Cuban women with college degrees are citizens in our sample and this statistic is roughly one fifth for high school graduates.

Central/South Americans have even lower percentages of 20 and 10, respectively. This may be one explanation for the lower median returns for married Mexican women.

5.3 Note on Endogeneity

As pointed out in Henderson et al. (2011), the econometric literature has yet to develop an approach which will adequately address endogeneity when the effects are heterogeneous. Specifically, they argue that we may need a unique instrument for each sub-group. This is potentially feasible, but even if we were able to find such a set of instruments, we would need to address potential heterogeneity within each sub-group. In other words, if the returns to education vary across individuals, the instrumental variable (IV) estimates give higher weight to those individuals more effected by the chosen instrument (Card, 2001; Imbens and Angrist, 1994). Thus, it may be the case that the IV results are more problematic than helpful (Harmon et al., 2003; Koop and Tobias, 2004). Harmon et al. (2003) further argue that ability bias and measurement error often cancel each other out; which leaves no advantage for an IV model. Ignoring these potential problems, we could also take the word of Card (2008) who notes that the existing evidence from IV estimations are often quite similar to the least-squares results. Given these potential problems, we leave the endogeneity question for future research.

6 Conclusion

In this paper, we used a partially linear additive nonparametric model to model to study the returns to schooling literature. Using the 2010 March CPS data, we estimated a Mincer style model. We estimated twelve separate regressions based on race, sex and martial status and found significant heterogeneity both within and between groups. Similar to past research, we found that blacks and Hispanics had higher returns on average. Different from past research, we were able to find some cases of uniformly higher returns. For example, in our full sample, we failed to reject the null of first-order stochastic dominance for the returns of both single and married Hispanic females over corresponding white females. That being said, this result depends upon the level of education. When we only looked at females whose highest level of education was a high-school diploma, we no longer found first-order dominance. However, that dominance showed up again when looking at females holding college degrees. In our male cases, we found higher returns on average for blacks and Hispanics, but found the highest returns predominantly belonged to whites. This shows the importance of looking at

heterogeneity and examining the entire distribution of gradient estimates.

References

- [1] Basu R., Ullah A. 1992. Chinese Earnings-Age Profile: A Nonparametric Analysis. *Journal of International Trade and Economic Development*. 1: 151-165.
- [2] Becker GS. 1964. *Human Capital, A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press: New York.
- [3] Belman D., Heywood J. 1991. Sheepskin effects in the return to education. *Review of Economics and Statistics*. 720-724.
- [4] Brand J.E., Xie Y. 2010. Who Benefits from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education. *American Sociological Review*. 75: 273-302.
- [5] Buchinsky M. 1998. The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*. 13: 1-30.
- [6] Buja A., Hastie T.J., Tibshirani R.J. 1989. Linear Smoothers and Additive Models. *Annals of Statistics*. 17: 453-555.
- [7] Cai Z. 2002. A Two-Stage Approach to Additive time-series Models. *Statistica Neerlandica*. 56: 415-433.
- [8] Cai Z., Fan J. 2000. Average Regression Surface for Dependent Data. *Journal of Multivariate Analysis*. 75: 112-142.
- [9] Cai Z., Masry E. 2000. Nonparametric Estimation of Additive Nonlinear ARX time-series: Local Linear Fitting and Projection. *Econometric Theory*. 16: 465-501.
- [10] Card D. 2008. Returns to Schooling. In S. N. Durlauf, L. E. Blume (Eds.). *The New Palgrave Dictionary of Economics Second Edition*. Palgrave Macmillan.
- [11] Card D. 1999. The Causal Effect of Education on Earnings. In O. Ashenfelter, D. Card (Eds.). *Handbook of Labor Economics*. Elsevier: Amsterdam.
- [12] Card D. 2001. Estimating The Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*. 69: 1127-1160.
- [13] Card D., Lemieux T. 2001. Can Falling Supply Explain the Rising Return to College for Younger Males? *Quarterly Journal of Economics*. 705-746.

- [14] Card D., Krueger A.B. 1992a. Does School Quality Matter: Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*. 100: 1-40.
- [15] Card D., Krueger A.B. 1992b. School Quality and Black-White Relative Earnings: A Direct Assessment. *Quarterly Journal of Economics*. 107: 151-200.
- [16] Duncan B., Hotz V. J., Trejo S. J. 2006. Hispanics in the U.S. Labor Market In Tienda M., Mitchell F. (Eds.). *Hispanics and Future of America*. The National Academies Press: Washington D.C.
- [17] Eren O., Henderson D.J. 2008. The Impact of Homework on Student Achievement. *Econometrics Journal*. 11: 326-348.
- [18] Griliches Z. 1977. Estimating the Returns to Schooling: Some Econometric Problems. *Econometrica*. 45: 1-22.
- [19] Harmon C., Hogan V., Walker I. 2003. Dispersion in the Economic Return to Schooling. *Labour Economics*. 10: 205-214.
- [20] Heckman J., Lochner L., Todd P. 2003. Fifty Years of Mincer Earnings Regressions. Technical Report 9732, National Bureau of Economic Research.
- [21] Heckman J., Lochner L., Todd P. 2006. Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E.A. Hanushek and F. Welch (Eds.). *Handbook of the Economics of Education*. 1: 307-458.
- [22] Heckman J., Lochner L., Todd P. 2008. Earnings Functions and Rates of Return. *Journal of Human Capital*. 2: 1-31.
- [23] Heckman J., Polachek S.W. 1974. Empirical Evidence on Functional Form of the Earnings Schooling Relationship. *Journal of the American Statistical Association*. 69: 350-354.
- [24] Henderson D.J., Maasoumi E. 2014. Searching for rehabilitation in nonparametric regression models with exogenous treatment assignment. In J.S. Racine, L. Su and A. Ullah (Eds.). *Oxford Handbook of Nonparametric and Semiparametric Econometrics and Statistics*. 501-520.

- [25] Henderson D.J., Parmeter C.F. 2014. *Applied Nonparametric Econometrics*. Cambridge University Press: New York.
- [26] Henderson D.J., Polachek S.W., Wang L. 2011. Heterogeneity in Schooling Rates of Return. *Economics of Education Review*. 30: 1202-1214.
- [27] Horowitz J.L. 2009. *Semiparametric and Nonparametric Methods in Econometrics*. Springer-Verlag: New York.
- [28] Imbens G., Angrist J. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*. 62: 467-476.
- [29] Kim W., Linton O.B., Hengartner N.W. 1999. A Computationally Efficient Estimator for Additive Nonparametric Regression with Bootstrap Confidence Intervals. *Journal of Computational and Graphical Statistics*. 8: 278-297.
- [30] King M., Tertilt M. 2003. IPUMS-CPS: An Integrated Version of the March Current Population Survey 1962-2002. *Historical Methods*. 36: 35-40.
- [31] King M., Ruggles S., Alexander J. T., Flood S., Genadek K., Schroeder M.B., Trampe B., Vick R. 2010. Integrated Use Microdata Series, Current Population Survey: Version 3.0. [Machine-Readable Database]. Minneapolis: University of Minnesota.
- [32] Koop G., Tobias J. 2004. Learning About Heterogeneity in Returns to Schooling. *Journal of Applied Econometrics*. 19: 827-849.
- [33] Landale N. S., Oropesa R. S., Bradatan C. 2006. Hispanic Families in the United States: Family Structure and Process in an Era of Family Change. In Tienda M., Mitchell F. (Eds.). *Hispanics and Future of America*. The National Academies Press: Washington D.C.
- [34] Lemieux T., Card D. 2001. Education, earnings, and the 'Canadian G.I. Bill'. *Canadian Journal of Economics*. 34: 313-344.
- [35] Li Q., Racine J. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press: Princeton.
- [36] Linton O. 1997. Efficient Estimation of Additive Nonparametric Regression Models. *Biometrika*. 84: 469-474.

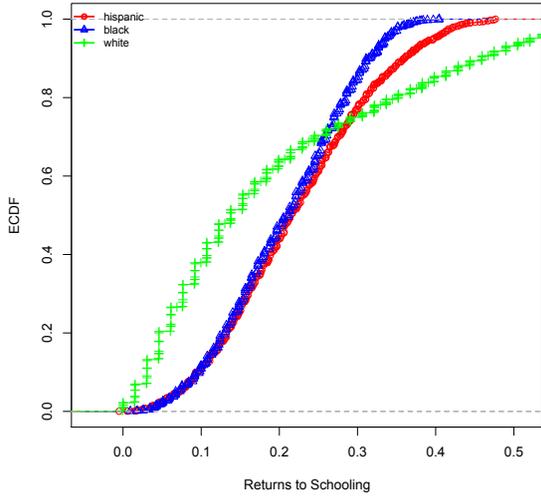
- [37] Linton O. 2000. Efficient Estimation of Generalized Additive Nonparametric Regression Models. *Econometric Theory*. 16: 502-523.
- [38] Linton O., Mammen E. 2008. Nonparametric Transformation to White Noise. *Journal of Econometrics*. 142: 241-264.
- [39] Linton O., Nielsen J. P. 1995. A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika*. 82: 93-100.
- [40] Mammen E., Linton O., Nielsen J. 1999. The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions. *Annals of Statistics*. 27: 1443-1490.
- [41] Mammen E., Park B.U., Schienle M. 2012. Additive Models: Extensions and Related Models. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*. L. Su, J. Racine and A. Ullah (Eds.), Oxford University Press: Oxford.
- [42] Manzan S., Zerom D. 2005. Kernel Estimation of a Partially Linear Additive Model. *Statistics and Probability Letters*. 72: 313-322.
- [43] Martins F.O.M. 2001. Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labor Force in Portugal. *Journal of Applied Econometrics*. 16: 23-39.
- [44] Martins-Filho C., Yang K. 2007. Finite Sample Performance of Kernel-Based Regression Methods for Nonparametric Additive Models under Common Bandwidth Selection Criterion. *Journal of Nonparametric Statistics*. 19: 23-62.
- [45] Masry E. 1996. Multivariate Local Polynomial Regression Estimation for time-series: Uniform Strong Consistency and Rates. *Journal of time-series Analysis*. 17: 571-599.
- [46] Masry E., Fan J. 1997. Local Polynomial Estimation of Regression Functions for Mixing Processes. *Scandinavian Journal of Statistics*. 24: 165-179.
- [47] Mincer J. 1974. *Schooling, Experience, and Earnings*. Columbia University Press: New York.
- [48] Murphy K.M., Welch F. 1990. Empirical Age-Earnings Profiles. *Journal of Labor Economics*. 8: 202-209

- [49] Newey W. 1994. Kernel Estimation of Partial Means. *Econometric Theory*. 10: 233-253.
- [50] Pagan A., Ullah A. 1999. *Nonparametric Econometrics*. Cambridge University Press.
- [51] Park J.H. 1994. Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment. Working Paper #338, Princeton University (Industrial Section).
- [52] Polachek S.W. 1975. Differences in Expected Post-School Investment and Determinant of Market Wage Differentials. *International Economic Review*. 16: 451-470.
- [53] Polachek S.W. 2008. Earnings Over the Life Cycle: The Mincer Earnings Function and Its Applications. *Foundations and Trends in Microeconomics*. 4: 165-272.
- [54] Robinson P.M. 1988. Root-n Consistent Semiparametric Regression. *Econometrica*. 56: 931-954.
- [55] Schick A. 1996. Root-n Consistent and Efficient Estimation in Semiparametric Additive Regression Models. *Statistics and Probability Letters*. 30: 45-51.
- [56] Severance-Lossin, E., Sperlich, S., 1999. "Estimation of derivatives for additive separable models," *Statistics*, 33, 241-265.
- [57] Schneider B., Martinez S., Owens A. 2006. Barriers to Educational Opportunities for Hispanics in the United States. In Tienda M., Mitchell F. (Eds.). *Hispanics and Future of America*. The National Academies Press: Washington D.C.
- [58] Spence, M. 1973. Job Market Signaling. *Quarterly Journal of Economics*. 87: 355-374.
- [59] Stock J.H., Watson M.W. 2007. *Introduction to Econometrics*. Addison Wesley: Boston.
- [60] Stone C. 1980. Optimal Rates of Convergence for Nonparametric Estimators. *Annals of Statistics*. 8: 1348-1360.
- [61] Stone C. 1985. Additive Regression and Other Nonparametric Models. *Annals of Statistics*. 13: 689-705.
- [62] Tjøstheim D., Auestad B. 1990. Identification of Nonlinear time-series: First-Order Characterization and Order Determination. *Biometrika*. 77: 669-687.

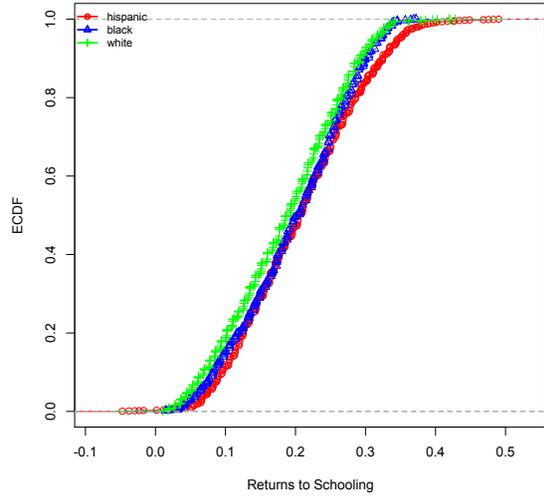
- [63] Tjøstheim D., Auestad B. 1994. Nonparametric Identification of Nonlinear time-series Projections. *Journal of the American Statistical Association*. 89: 1389-1409.
- [64] Ullah, A. 1985. Specification Analysis of Econometric Models. *Journal of Quantitative Economics*. 1: 187-209.
- [65] Welch F. 1973. Black and White Differences in Returns to Schooling. *American Economic Review*. 63: 893-907.
- [66] Xiao Z., Linton O.B., Carroll R.J., Mammen E. 2003. More Efficient Local Polynomial Estimation in Nonparametric Regression With Autocorrelated Errors. *Journal of the American Statistical Association*. 98: 980-992.

Yang, L., Sperlich, S., Hardle, W. 2003. "Derivative estimation and testing in generalized additive models," *Journal of Statistical Planning and Inference*, 115, 521-542.

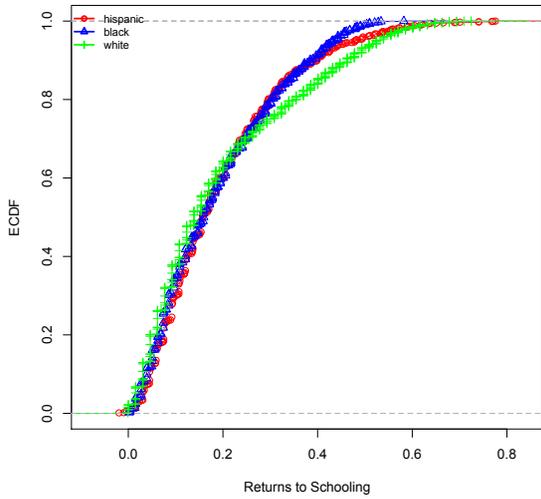
Figure 1: Returns to schooling for the pooled samples



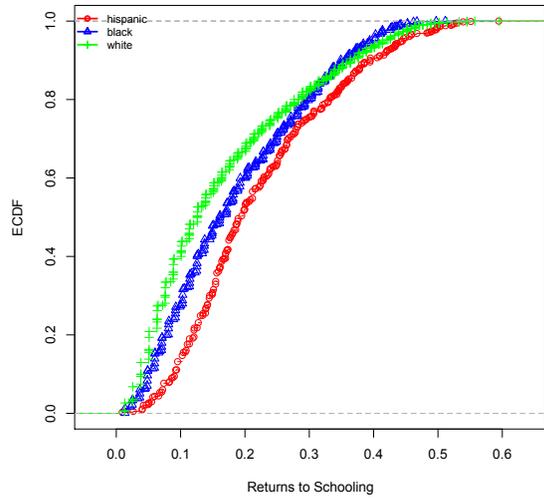
(a) Married Male



(b) Married Female

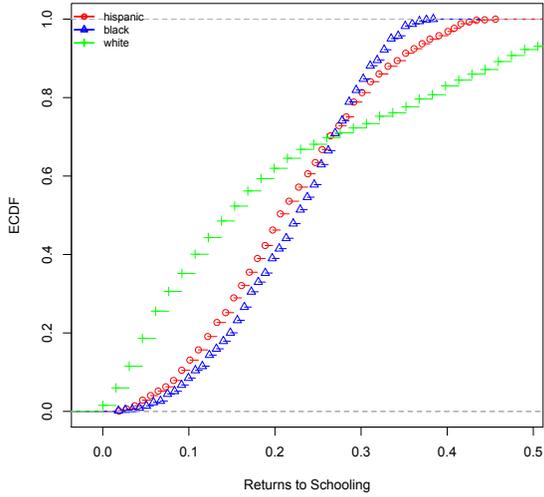


(c) Single Male

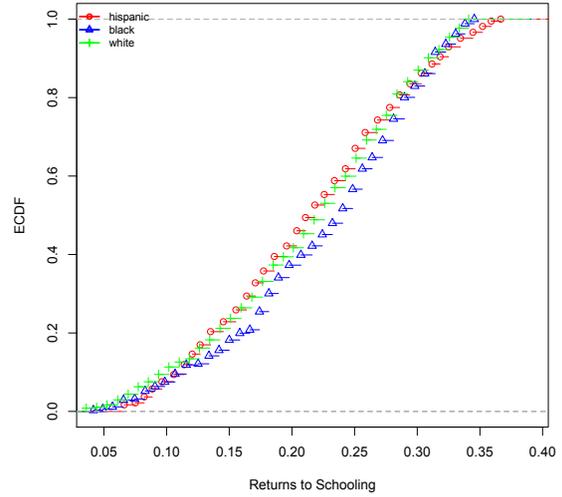


(d) Single Female

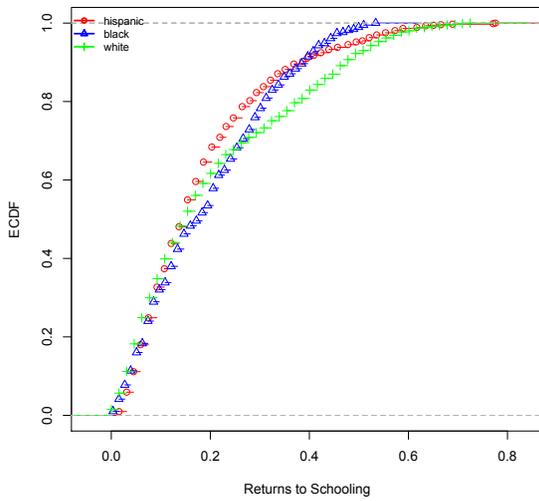
Figure 2: Returns to schooling for high school graduates



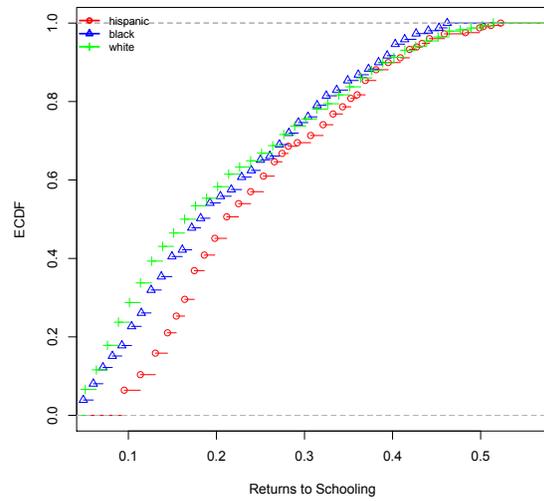
(a) Married Male



(b) Married Female

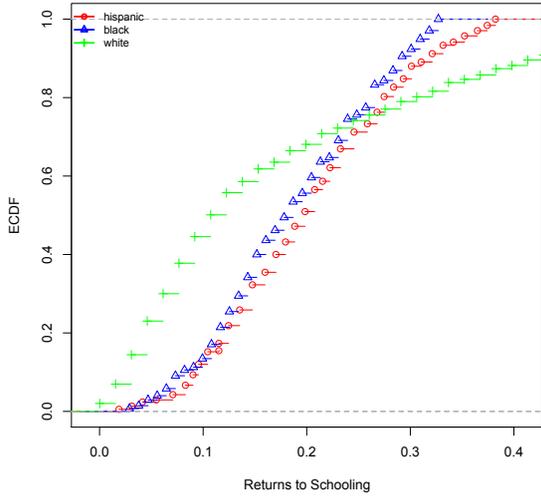


(c) Single Male

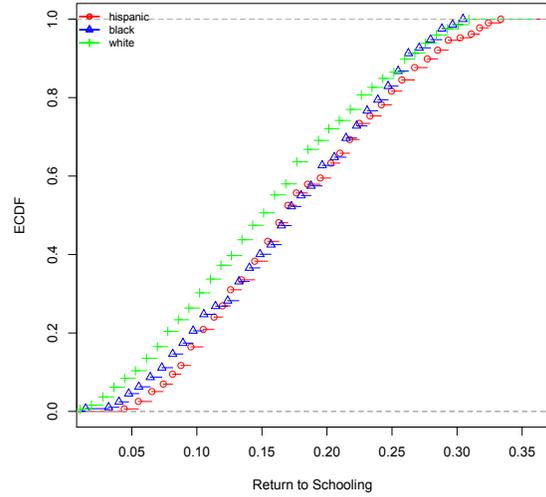


(d) Single Female

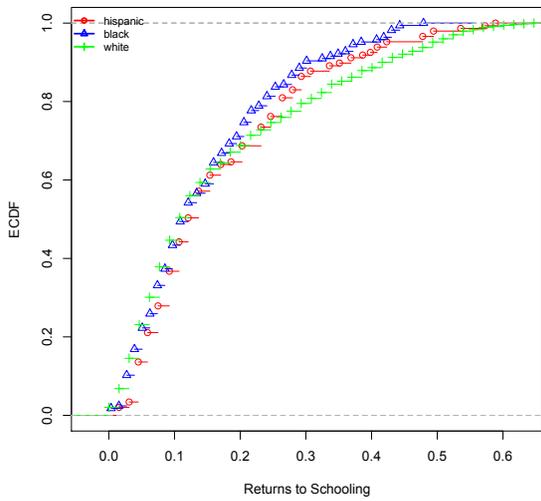
Figure 3: Returns to schooling for college graduates



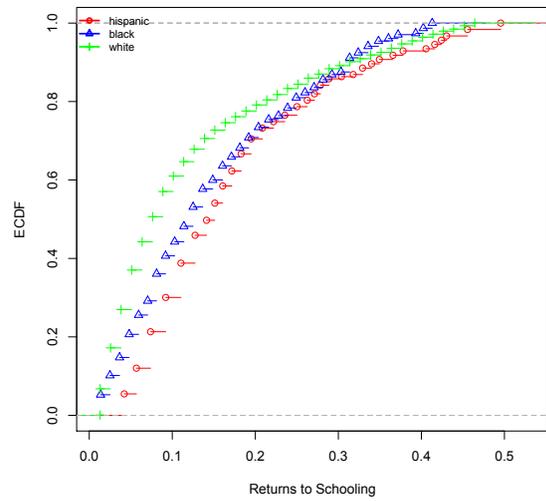
(a) Married Male



(b) Married Female



(c) Single Male



(d) Single Female

Table 1: Descriptive statistics for our sample. Median and mean values are reported in the table along with the standard deviation and sample size for each sub-group.

Variable	Married Male			Married Female			Single Male			Single Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
<i>Schooling</i>												
Median	14	13	12	14	13	12	13	13	12	14	13	13
Mean	14.41	13.71	12.03	14.28	14.00	12.53	13.77	13.28	11.97	14.63	13.77	12.90
Std Dev	2.51	2.25	3.14	2.35	2.33	3.31	2.25	2.12	2.67	2.30	2.18	2.82
<i>Experience</i>												
Median	24	26	22	23	24	22	9	13	10	9	15	12
Mean	24.02	25.34	23.01	23.04	23.18	21.89	12.72	15.43	11.95	12.67	16.54	14.44
Std Dev	10.19	10.13	10.28	9.84	9.67	10.02	10.66	10.93	9.04	9.86	10.28	9.898
<i>ln (Earnings)</i>												
Median	2.97	2.74	2.59	2.70	2.57	2.42	2.57	2.42	2.27	2.54	2.44	2.34
Mean	3.00	2.75	2.61	2.68	2.62	2.43	2.58	2.47	2.34	2.57	2.47	2.37
Std Dev	0.64	0.62	0.63	0.59	0.58	0.59	0.64	0.63	0.63	0.60	0.58	0.58
<i>n</i>	16941	1576	3293	10869	1233	1981	4337	962	1757	2946	1323	1017

Table 2: Median return to schooling estimate across the sub-group for the parametric models. Heteroskedasticity robust standard errors are listed below each median estimate.

Model	Married Male			Married Female			Single Male			Single Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
1	0.1029	0.1035	0.0776	0.1144	0.1121	0.0893	0.0966	0.1305	0.0701	0.1159	0.1200	0.0851
	0.0018	0.0066	0.0034	0.0022	0.0069	0.0043	0.0040	0.0088	0.0057	0.0044	0.0068	0.0068
2	0.1038	0.1051	0.0785	0.1155	0.1125	0.0922	0.0978	0.1310	0.0704	0.1185	0.1253	0.0901
	0.0018	0.0066	0.0034	0.0022	0.0069	0.0044	0.0040	0.0088	0.0057	0.0045	0.0070	0.0069
3	0.1038	0.1050	0.0786	0.1155	0.1130	0.0921	0.0978	0.1312	0.0706	0.1190	0.1253	0.0903
	0.0018	0.0066	0.0035	0.0022	0.0069	0.0044	0.0040	0.0088	0.0057	0.0045	0.0070	0.0069
4	0.1014	0.1055	0.0767	0.1150	0.1031	0.0917	0.0943	0.1198	0.0713	0.1127	0.1161	0.0999
	0.0062	0.0006	0.0350	0.0011	0.0148	0.0349	0.0055	0.0198	0.0404	0.0139	0.0142	0.0371
5	0.3718	0.5718	0.4292	0.1864	0.7610	0.1584	0.2919	0.1803	0.6105	0.3544	0.3709	0.2355
	0.1031	0.1659	0.2146	0.0247	0.2506	0.0670	0.0764	0.0397	0.2711	0.0939	0.0973	0.0862
6	0.7923	-1.2653	1.8828	0.4300	0.7997	-0.0913	0.0502	1.4526	1.3542	0.3354	-0.6859	-0.1410
	0.1772	0.7827	1.6624	0.0968	0.2469	0.1557	0.4156	1.1338	0.7196	0.0856	0.7507	0.3510

Table 3: Median return to schooling estimate across the sub-group for the nonparametric specification. Heteroskedasticity robust (wild bootstrapped) standard errors are listed below each median estimate.

	Married Male			Married Female			Single Male			Single Female		
	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic	White	Black	Hispanic
Pooled	0.1378	0.2125	0.2163	0.1900	0.2060	0.2053	0.1387	0.1594	0.1569	0.1186	0.1719	0.1949
	0.0025	0.0051	0.0041	0.0041	0.0308	0.0041	0.0094	0.0071	0.0064	0.0062	0.0056	0.0844
HS	0.21531	0.2293	0.2062	0.2262	0.2406	0.2183	0.1542	0.1833	0.1540	0.1700	0.1821	0.2117
	0.0016	0.0054	0.0095	0.0076	0.0778	0.0074	0.0004	0.0081	0.00054	0.0011	0.0125	0.0114
College	0.1072	0.1809	0.1892	0.1512	0.1633	0.1666	0.1079	0.1209	0.1208	0.0762	0.1251	0.1512
	0.0015	0.0044	0.0091	0.0071	0.0764	0.0068	0.0003	0.0054	0.0047	0.0011	0.0115	0.0106

Table 4: Descriptive statistics for Hispanic females. Median and mean values of schooling are reported in the table along with the sample size for geographic origin.

Schooling	Married Female			Single Female		
	<i>n</i>	Median	Mean	<i>n</i>	Median	Mean
Mexico	1198	12	11.95	524	13	12.71
Puerto Rico	181	14	14.1	141	13	13.51
Cuba	77	14	14.48	36	14	14.16
Central/South America	396	13	12.79	243	12	12.62
Pooled	1981	12	12.53	1017	13	12.90

Table 5: Hispanic Females - Median return to schooling estimate across the sub-group for the nonparametric specification. Heteroskedasticity robust (wild bootstrapped) standard errors are listed below each median estimate.

Schooling	Married Female				Single Female			
	Mexico	Puerto Rico	C/S America	Cuba	Mexico	Puerto Rico	C/S America	Cuba
Pooled	0.2030 0.0088	0.2026 0.0008	0.2124 0.0086	0.2062 0.0076	0.2236 0.0118	0.1830 0.0012	0.1887 0.0011	0.2085 0.0114
High School Degree	0.1958 0.0076	0.2340 0.0067	0.2339 0.0078	0.2594 0.0078	0.2251 0.0111	0.2049 0.0111	0.2117 0.0988	0.1071 0.0107
College Degree	0.1630 0.0072	0.1732 0.0071	0.1630 0.0070	0.2100 0.0073	0.1512 0.0120	0.0937 0.0619	0.1085 0.0937	0.0619 0.0109