Heterogeneity in Schooling Rates of Return

Daniel J. Henderson[*]
State University of New York at Binghamton and IZA

Solomon W. Polachek
State University of New York at Binghamton and IZA

Le Wang
University of New Hampshire

May 8, 2009

Abstract

This paper relaxes the assumption of homogenous rates of return to schooling by employing nonparametric kernel regression. This approach allows us to examine the differences in rates of return of education both across and within groups. Similar to previous studies we find that on average blacks have higher returns to education than whites, natives have higher returns than immigrants and younger workers have higher returns than older workers. Contrary to previous studies we find that the average gap of the rate of return between white and black workers is larger than previously thought and the gap is smaller between immigrants and natives. We also uncover significant heterogeneity both across and within groups. The estimated densities of returns vary across groups and time periods and is often skewed. Further examination of the densities shows significant portions of the population possessing negative returns. The negative returns are more common in older workers, individuals with young children and immigrants.

Keywords: Mincer Regressions, Nonparametric Kernel
JEL Classification: C14, J24

---

[*]Daniel J. Henderson, Department of Economics, State University of New York, Binghamton, NY 13902-6000, (607) 777-4480, Fax: (607) 777-2681, e-mail: djhender@binghamton.edu.

1   Introduction

In regression analysis it is often argued that individuals with different backgrounds not only differ in their intercepts, but also in their reactions to changes in explanatory variables. The simplest way to model this type of parameter heterogeneity is to include both an additive and interaction dummy variable into the regression. This allows for both differences in intercepts across groups and different slope parameters.

The rate of return of schooling literature is not ignorant of this development, however, most of the literature assumes that all individuals within a particular group have the same return to a one-unit increase in education.[1] There are good reasons to believe that rate of return to schooling varies within a particular sub-group. It is also natural to believe that this variation may differ across groups and over time. Uncovering this variability may contain useful information for policy makers. Panel data estimators exist which will relax these assumptions, but they require repeated observations across individuals. This information is not always readily available. Random coefficients models allow for parameter heterogeneity and estimation in a cross-section, but are subject to the same functional form restrictions as typical panel data estimators (Harmon, Hogan and Walker 2003). Specifically, if the specified parametric functional form is incorrect, estimation generally leads to inconsistent estimates. Further, these methods often require an assumption regarding the distribution of the variability. On the other hand, nonparametric regression methods allow for separate rate of return estimates with respect to each realization of the regressors. They also eschew functional form assumptions and are consistent under a more broad range of data generating processes (Heckman, Lochner and Todd 2008). In this paper we use both parametric and nonparametric methods to investigate heterogeneity across groups. We further use the nonparametric methods to investigate heterogeneity within groups.

Our parametric results are generally consistent with the literature. We find that the rate of return has been rising over time, albeit non-monotonically. The rate of return to older individuals is less than that of younger workers. The return black workers receive is greater than their white counterparts. We also find that the rate of return is higher for natives than immigrants. Following cohorts over time we find

---

[1] Although it can be argued that schooling and education are distinct, we speak of them as being synonymous in this paper.

that the rate of return falls with age, but rises within a particular age-group over time. We also find that the drop in the rate of return over time is faster for blacks than for whites. A similar result is found for native versus immigrant workers.

The results from the nonparametric regressions on average are often in line with the parametric findings. We find that the rate of return of education (on average) is larger for younger as opposed to older individuals. We also find that blacks have a higher rate of return than whites and that natives have higher rates of return of schooling than immigrants. Although the direction of inequalities at the median generally remain the same between the two methods, the magnitude of the differences often do not. For instance, when comparing the parametric and nonparametric results, we notice that for blacks, the parametric estimates are generally smaller than their nonparametric counterparts, indicating that the parametric estimates are downward biased. Although the magnitude of the median partial effect is also higher in the nonparametric model for whites, this difference is lesser and hence the gap between the two groups is larger when employing nonparametric regression.

Perhaps more important than differences at the median is the uncovering of substantial heterogeneity in the rates of return of schooling within groups. As stated previously, many papers allow for heterogeneity across groups, but few allow for heterogeneity within specific groups. We show substantial variation in the rate of return within groups. However, this variation is not consistent across groups or across time. For example, we show that the variation in the rate of return estimates is generally much larger for natives than for immigrants. Also, we find a (non-monotonic) decrease in the variation of the rate of return for black workers over time.

We find one striking unexpected result when examining heterogeneity in the rates of return. This is the presence of negative rates of return. We find significant percentages of negative returns for specific groups. For example, for workers aged 40 and over, we always have double digit percentages of negative rates of return. Further examination shows that these percentages increase with age. We also find that whites are generally more likely to have negative returns than blacks, and until 1990, immigrants were more likely to have negative returns than natives. A further breakdown of the results by other characteristics show which characteristics (e.g., having a child under the age of five) makes it more likely to have negative rates of return.

The remainder of the paper is organized as follows: Section 2 describes both the parametric and nonparametric estimation methods. The third section discusses the data whereas the fourth section gives the results for the parametric models. Section 5 gives the results for the nonparametric estimates and the sixth section discusses future research directions. Finally, the seventh section concludes.

2 Empirical methodology

In regression we are typically concerned with predicting the left-hand-side variable given specific values of one or more right-hand-side variables. Mathematically we note that for a particular observation, this is the conditional expectation $E\left(y_i|x_i = x\right)$. The general regression model with an additive mean zero random error is written as

$$y_i = E\left(y_i|x_i\right) + u_i.$$

We often forget this step and assume that $E\left(y_i|x_i\right)$ is linear in x, i.e. $E\left(y_i|x_i\right) = \alpha + \beta x_i$. If this model is true and the other Gauss-Markov assumptions hold, then the estimators of α and β are the best linear unbiased estimators and we can proceed with inference and policy suggestions. However, if the true model is nonlinear and we ignore this, estimation may not only lead to inconsistent estimates, but it can also mask important heterogeneity in the marginal effects. For example, suppose the true model is quadratic in x, but we fit a linear model. In a linear model the estimated partial effect $\partial y/\partial x = \beta$ is constant for all x. Thus, not only will the linear model's result be inconsistent, but it will ignore that the true partial effect varies with x. Even worse, the marginal effect could take both positive and negative values. If we were to implement policy by using results from the linear model, but the technology was quadratic, this could lead to unintended consequences for a particular group, for example, a detrimental instead of positive impact of a treatment for a sub-group of the population.

Given that we generally do not know the true data generating process, we have a few options: (1) Simply hope that the true model is linear. Given that this is only one possibility out of an infinite number of possibilities, this may be a bit naive. (2) Fit higher order polynomials as well as use interaction terms. This is a promising approach, but given the number of possibilities, it is difficult to model all of these

without quickly running out of degrees of freedom. (3) Let the data tell the form of the technology. This is the approach we take.

2.1 Ordinary least squares

Although OLS is well understood by economists, we feel the need to briefly describe the estimator in this particular framework. The typical Mincer (1974) regression model is given by

$$\ln y_i = \alpha + \beta s_i + \gamma t_i + \delta t_i^2 + u_i, \qquad i = 1, ..., n \qquad (1)$$

where y is annual earnings, s is the number of years of education and t represents the years of experience. α, β, γ and δ are parameters to be estimated and u is the standard additive error term. Note that even though our left-hand-side variable is measured in logs and experience enters in quadratic form, this model is linear in its parameters and thus may be estimated by OLS.

Of particular interest in this paper is the coefficient attached to the schooling variable, β. It represents the partial change in $\ln y$ when s is changed by one unit. It is roughly interpreted as the percentage change in earnings when schooling is increased by one year. This value is fixed for all levels of schooling which means that it is assumed that a one year increase in schooling brings about the same percentage change in earnings regardless of the number of years of schooling. Further, the model assumes that schooling is linearly separable from the other regressor(s) in the model. In other words, this coefficient is assumed constant across groups/individuals. Further, this model is subject to the same criticism as other parametric models. Specifically, misspecification of the conditional mean will generally lead to inconsistent parameter estimates and potentially inappropriate policy prescriptions. Although we are aware that more sophisticated versions of (1) are available in the literature (Heckman and Polachek 1974; Heckman, Lochner and Todd 2008), this form is still the standard when analyzing the rate of return of education.

2.2 Generalized kernel estimation

The basic idea behind nonparametric regression is to estimate the conditional mean arbitrarily close given a sufficient amount of data. Here we consider a variant of the

local-linear least-squares (LLLS) estimator (Fan and Gijbels 1992; Pagan and Ullah 1999). In short, LLLS performs weighted least-squares regressions around a point x with weights determined by a kernel function and bandwidth vector. Specifically, more weight is given to observations in the neighborhood of x. This is performed over the range of x and then the unknown function is estimated by connecting the point estimates. Some of the benefits of LLLS are that it requires no functional form assumptions on the underlying functional form and allows for heterogeneity in the partial effects. Further, if indeed the true functional form is linear, the LLLS estimator nests the OLS estimator when the bandwidth is allowed to be very large.

Specifically, we use Generalized Kernel Estimation (Li and Racine 2004; Racine and Li 2004) to estimate the conditional mean and gradient. To begin, first consider the nonparametric regression model

$$y_i = m(x_i) + \varepsilon_i, \qquad i = 1, ..., n \tag{2}$$

where y_i is the dependent variable measured for observation i. $m(\cdot)$ is the unknown smooth function with argument $x_i = [x_i^c, x_i^u, x_i^o]$, x_i^c is a vector of continuous regressors, x_i^u is a vector of regressors that assume unordered discrete values, x_i^o is a vector of regressors that assume ordered discrete values, ε is an additive error, and n is the number of observations. In our application, y is log annual earnings and x^c contains $q = 2$ elements: years of education and experience. x^u contains a single element for whether or not the individual was top-coded and there are no elements in x^o.

Taking a first-order Taylor expansion of (2) with respect to x yields

$$y_i \approx m(x) + (x_i^c - x^c)\beta(x) + \varepsilon_i \tag{3}$$

where $\beta(x)$ is defined as the partial derivative of $m(x)$ with respect to x^c. The LLLS estimator of $\delta(x) \equiv (m(x), \beta(x))'$ is given by

$$\widehat{\delta}(x) = (X'K(x)X)^{-1} X'K(x)y, \tag{4}$$

where X is a $n \times (q+1)$ matrix with ith row being $(1, (x_i^c - x^c))$ and $K(x)$ is a diagonal n matrix of kernel weighting functions for mixed continuous and categorical data with bandwidth parameter vector h (Li and Racine 2007). Closer inspection of the estimator in (4) shows that the estimate is specific to x. In other words, we obtain a fitted value and derivative estimate (for each regressor) for each x. This allows us to observe heterogeneity in the partial effect of schooling.

2.2.1 Bandwidth selection

Estimation of the bandwidths (h) is typically the most salient factor when performing nonparametric regression. For example, choosing a very small bandwidth means that there may not be enough points for smoothing and thus we may get an undersmoothed estimate (low bias, high variance). On the other hand, choosing a very large bandwidth, we may include too many points and thus get an oversmoothed estimate (high bias, low variance). This trade-off is a well-known dilemma in applied nonparametric econometrics and thus we usually resort to automatic determination procedures to estimate the bandwidths. Although there exist many selection methods, one increasingly popular method is Hurvich, Simonff and Tsai's (1998) corrected expected Kullback Leibler (AIC_c) criterion. This method chooses smoothing parameters using an improved (in terms of bias) Akaike Information Criterion. Specifically, the bandwidths are chosen to minimize

$$AIC_c(h) = \log\left(\widehat{\sigma}^2\right) + \frac{1 + tr(H)/n}{1 - [tr(H) + 2]/n}, \tag{5}$$

where

$$\begin{aligned} \widehat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{m}(x_i)\right)^2 \\ &= \left(\frac{1}{n}\right)y'(I - H)'(I - H)y, \end{aligned} \tag{6}$$

where I is an identity matrix of dimension n and $\widehat{m}(x_i) = Hy_i$. H is a linear smoother matrix composed of kernel weights and depends on x but not on y. The first term in (5) rewards fit while the second term is the penalty function.

2.2.2 Estimation of the density of the partial effects

One benefit of nonparametric kernel methods is that they give a plethora of results. Observation specific estimates can be obtained for each regressor in a local-linear regression. But the problem with $n \times q$ estimates is that one has $n \times q$ estimates. It is often difficult and/or impractical to present this many values in a paper. Therefore researchers often devise ways to present the results. Some authors simply look at the mean or median of the estimates for a particular regressor. However, this ignores possible heterogeneity in the estimates. Others attempt to look at the quartiles and/or

percentiles, but these also may hide some interesting findings. One increasingly popular method to present the results is to plot kernel densities of the estimates. This allows us to examine the entire set of estimates for a particular regressor in one simple-to-view figure.

Plotting kernel densities of predicted values and/or derivatives is analogous to that for a simple vector of data. Let $\widehat{\beta}_i = \widehat{\beta}(x_i)$, then the kernel density estimate for the estimated derivative is defined as

$$\widehat{f}\left(\widehat{\beta}\right) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{\widehat{\beta}_i - \widehat{\beta}}{h}\right), \qquad (7)$$

where h is a scalar bandwidth and $K\left(\cdot\right)$ is the kernel function. Stated loosely, a kernel density estimate can be thought of as a smoothed histogram. The kernel function determines the shape of the bumps and the bandwidth controls the degree of smoothness.[2]

3 Data

The data are obtained from the Integrated Public Use Microdata Series (IPUMS) 1940 - 2000 and the American Community Survey (ACS) 2005. IPUMS data 1940 - 2000 are based on U.S. Decennial Censuses.[3] For 1940 - 1970, the 1% samples (1-in-100 national random sample of the population) are available; for 1980 - 2000, the 5% samples (1-in-20 national random sample of the population) are available. Since Census data are conducted every ten years, we also pool together ACS 2005 (1-in-100 national random sample of the population) for our analysis to reflect recent trends.[4] Our sample is restricted to male workers aged 16 and above. We focus on individuals from "regular households" and "additional households under the 2000 definition". Individuals living in group quarters are excluded from the sample. Moreover, we keep only the observations with unaltered or logically edited values. The observations that are manually edited using hotdeck, colddeck, or the unspecified allocation method are excluded from our analysis.

---

[2]Throughout the paper, when presenting kernel density estimates, we employ a Gaussian kernel and use the Silverman (1986) rule-of-thumb bandwidth.

[3]All questions were asked for the previous calendar year. For instance, reported earnings in the 1980 census are earnings in 1979.

[4]All the data files can be downloaded on http://usa.ipums.org/.

To perform the analysis, three variables are of primary interest: individual earnings, years of schooling and working experience. The definition of individual total income (the sum of all sources of earnings) is different across years, as more detailed income categories are asked in later years. The only income category that is comparably defined across all years in our analysis is wage and salary income. Ideally, to separate income effects from labor supply effects, hourly wage is needed. Information on weeks and hours worked is, however, not consistently available. The variable "weeks worked in the previous year" is continuous in all years except in 1960 and 1970 where it is documented in intervals. The variable "hours usually worked per week in the previous year" is not available before 1980. Therefore, we construct the measure of individual income based on individual's annual wage and salary income.[5] To ensure comparability across years, we also adjust the wage for inflation using the adjustment factors provided by IPUMS.[6] Individuals with negative income are excluded. Moreover, the wage variable is top-coded with different topcodes across years. Therefore, we create a dummy variable, equal to one if income exceeds the topcodes and zero otherwise, that is included in the analysis whenever appropriate.[7]

Another issue is that schooling is not measured in a consistent fashion over time. Prior to 1990, the census asked individuals about how many grades of school or years of college they had. Starting from 1990, however, the census and ACS asked individuals about the highest grade or diploma completed. Moreover, grades completed below 9th grade are reported in three-year intervals. For data from 1990-2005, we replace these intervals with midpoints. The coding scheme utilized to convert intervals and degrees obtained to years of schooling is in the data appendix. Following the literature, our definition of potential working experience equals current age minus years of schooling minus 6. Negative values are recoded as zero.

Finally, a number of other variables will be used to further investigate potential heterogeneity in the return of education both across and within groups. These variables include a dummy variable indicating whether or not one is an immigrant, a dummy variable indicating self-employment status, region dummy variables (Mid-

---

[5]Any values exceeding 999999 are recoded as missing values.

[6]See IPUMS website for consumer price index adjustment factor for each year http://usa.ipums.org/usa-action/variableDescription.do?mnemonic=INCTOT.

[7]See IPUMS wesbiste for topcodes for each year http://usa.ipums.org/usa-action/codes.do?mnemonic=INCWAGE.

west, South, West, and East), a dummy variable indicating whether or not there are any children younger than five in the household, a dummy variable indicating if an individual is currently married, and dummy variables indicating whether an individual is white or if an individual is black.

4   Parametric estimates

Tables 1 and 2 present our parametric results of rates of return of education utilizing OLS estimation. Table 1 gives the results over time both for the pooled sample and for specific groups. For all regression estimates, robust standard errors are reported in parenthesis. Table 2 gives the breakdown by cohort results for the same groups in the previous table. This section is useful not only for comparison to the nonparametric results below, but also relative to that found in the literature.

4.1   Pooled results

We first present the pooled results separately for 1940 to 2005 in the first column of Table 1. Regression estimates imply that there exists a undoubtedly positive return of education in labor market, ranging from 9% to 13%. The results reflect the long-run trend in rates of return of education. Comparing the estimates from 1950 with those from 2005, the results indicate there is an increase in the return of education of 3% in the past fifty years. The increase was, however, not monotonic over time; the return to an additional year of schooling increased from 9% in 1950 to 10.4% in 1970, then fell to 9.5% in 1980. Afterwards, rate of return started to increase in 1990, stagnated in 2000, and continued to increase in 2005. These trends are more or less consistent with the literature (Goldin and Katz 1999; Card and Lemieux 2000).

To those unfamiliar with the literature, the result that stands out is that the rate of return fell in 1980. This can be partly explained by at least two phenomenon. First, there was a sharp increase in the supply of educated workers during that time period (Angrist and Chen 2007). Second, the recessions in the 1970's and early 1980's likely decreased the demand for labor. However, there is no reason to believe that a recession will affect younger and older workers equally. Hence, we deem it necessary to look for potential heterogeneity in the rate of return of education.

4.2 Sample splits

It is native to believe that all groups of individuals receive the same rate of return of education. Here we take the data used for the pooled sample and split it based on groups of interest. Specifically, we consider four age groups, two different groups based on race as well as splits by marital status and whether or not the individual was born in the United States. Specifically, we take the sample and split it according to one of the aforementioned categories. We then run OLS on each group for each year and report the results in the corresponding element of Table 1.

4.2.1 Age groups

The second through fifth rows of Table 1 present the results separately for four age groups (20-29, 30-39, 40-49, and 50-59). For all the groups, we again find a large, positive return of education, that is both economically and statistically significant. The estimates vary from about 6% to about 19%. The results also confirm that there has been a long-run increasing trend in the return of education, although such a trend is not monotonic. The return of education is higher for younger workers than older workers, and there is a monotonic decline in the return of education as age increases. For example, in 1940, the return was as large as 18.7% for workers aged 20-29 and fell to 14.3% for workers aged 30-39, to 10.7% for workers aged 40-49, and finally to 8.1% for workers aged 50-59. Similar patterns are also found for other years. Another feature is that the gap in the return of education between younger and older workers has been widening over time since 1950. For example, the difference in the return of education between workers aged 20-29 and workers aged 50-59 increased from about 4% in 1950 to roughly 11% in 2005.

One point of interest is the discussed drop in rates of return in 1980. We argued that the drop can partially be explained by both an influx of educated labor and the relatively weak economy. Note here that this effect was strongest for the youngest group (20-29). This group of workers were also most affected by the draft (Angrist and Chen 2007).This result is expected as they have the least experience to fall back upon. At the same time there was relatively little change in the next two groups (30-39 and 40-49), but a significant drop for older workers (50-59). This result is likely related with retirement trends (Soldo, Mitchell, Tfaily, and McCabe 2006; Clark and

Mitchell 2005).

4.2.2 Race

The results separately for white and black workers are presented in the next two columns of Table 1. While we continue to find a positive return of education for both groups, years of schooling are rewarded differently for white and black workers. Except in 1940, black workers consistently enjoy larger benefits from education than do their white counterparts (Welch 1973; Card and Krueger 1992). Another point is noteworthy. We continue to find a long-run, although not monotonic, increasing trend of the return of education for white workers. The pattern for black workers is, however, not consistent with the pooled results above. In particular, not only do we find a long-run increasing trend in the return of education; such a trend is also monotonic, increasing from about 8.9% in 1950, to about 11% in 1980, and to 16% in 2005. Moreover, this increase for black workers is also larger than that for white ones. Despite the increasing trend for black workers, there existed a gap in the return of education between black and white workers. The gap tended to converge before 1970, but started to diverge after 1970. For example, the difference in return of education between black and white workers was 1% in 1950, and decreased to 0.1% in 1990, but raised to about 4% in 2005.

4.2.3 Marital status

Estimates of the return of education by marital status are presented in the eighth and ninth columns of Table 1. The results for married individuals are broadly consistent with the non-monotonic, increasing trend above. However, the results for individuals who are not currently married are slightly different. Compared to the results in 1950, the results for later years indicate that there is an increase of at least 4% for individuals who are currently not married. For example, the pattern is consistent with the pooled results before 2000; the return of education first increased, fell slightly in 1980 and finally increased in 1990. The return of education, however, started to decline slightly in 2000 and stagnated in 2005. Non-married workers have a higher return of education than married workers. The gap is persistent over time, with the difference ranging from 2% to 8.6%.

4.2.4 Immigration status

The results by immigration status are presented in final two columns of Table 1. The results indicate that the return of education is considerably lower for immigrants than for natives (Chiswick 1978; Chiswick and Miller 2004; Rivera-Batiz 2007). For example, in 1950, while the return of education for natives is as large as 10%, it is only 3% for immigrants. However, while there is an increasing trend for both immigrants and natives, the growth in the return of education for immigrants is much faster than that for natives before 1990, closing the gap to some extent over time. In particular, the difference was about 7% in 1950, but fell to 2.4% in 1980. After 1990, the growth slowed down for immigrants but rose for natives, leading to a larger gap of 5.3% in 2005.

4.3 Cohort results

Comparison between age groups in a particular year may not necessarily reflect the life cycle pattern for returns to education. While we do not have panel data that allow us to follow the same individuals over time, a better understanding of the life cycle pattern can be obtained by tracing the return of education for the same cohort over time (Card and Lemieux 2000). The results for cohorts by group are presented in Table 2.

4.3.1 Race

The results by race are presented in the first two panels. First, we notice that the cross-sectional estimates (i.e. the results for different age groups in the same year) confirm that younger workers have higher returns to education than older ones, consonant with the results above. Also, comparing the results for black workers with those for white workers, we again find that education is more rewarding for black workers than for white workers.

Second, looking at the return of education for the same cohort over time (diagonal numbers), we find that the return of education for the same cohort is, in general, higher in their 20's than in their later lives. Such a result holds for both black and white men. The exceptions are mostly for those in 1980 when a decline in the return

of education occurred. Such exceptions are masked by comparing the age groups in the same year, and thus highlight the importance of looking at the cohort effects.

The decline in the return of education over the life cycle is usually larger for blacks than for whites. While the difference between black workers in their 20's and the same cohort in their 50's is around 6%, the difference between white workers in their 20's and the same cohort in their 50's is roughly 4%.

4.3.2 Marital status

The results by marital status are presented in the third and fourth panels of Table 2. First, looking at each column, cross-sectional estimates again show that the return of education declines as one gets older. Comparing the results for married workers with those who are not currently married, the return of education is in general larger for non-married workers than for married workers.

Second, consistent with the cross-sectional estimates, cohort-based estimates indicate a decline in the return of education over the life cycle. However, there are exceptions, especially for those in 1980 when the decline in the return of education occurred. The discrepancies between cross-sectional and cohort estimates again indicate that cross-sectional estimates fail to provide a complete picture of the return of education over the life cycle. Moreover, starting with a lower return of education, individuals aged 20-29 in 1980 experienced less of a decline in terms of changes over the life cycle. Comparing married individuals aged 20-29 in 1980 with the same cohort in 2005, the difference is only 0.6%.

We note here that we should be careful with the results found in this split. Even noting that these are random draws for a particular year, here we are not necessarily comparing the same type of individuals. Recall that our definition of married is whether or not the individual is married at the time of the census. For example, individuals who are in their 20's and married in 1990 are not necessarily those who are in their 30's and married in 2000. It is likely that some may be divorced, separated or widowed.

4.3.3　Immigration status

The results by immigration status are presented in final two panels of Table 2. First, both the cross-sectional and cohort-based estimates indicate a decrease in the return of education over the life cycle. However, the cross-sectional approach seems to overestimate (underestimate) the magnitudes of the decline for immigrants (natives). In 1940, for instance, the difference between immigrants aged 20-29 and those aged 30-39 is 10%; in contrast, the difference between immigrants aged 20-29 in 1940 and the same group aged 30-39 in 1950 is 7.8%. Second, consistent with the results above, the estimates imply that the return of education is in general lower for immigrants than for natives. Third, the results indicate that the decline in the return of education is smaller for immigrants than for natives. For instance, for immigrants aged 20-29 in 1960, the return of education fell from 8.5% to 7.2%, a decrease of 1.3%; for natives aged 20-29 in 1960, the return of education, on the other hand, dropped from 13.1% to 6.7%, a decrease of 6.4%.

5　Nonparametric estimates

Tables 3-6 present our nonparametric estimates of the rates of return of education. Table 3 gives the median of the nonparametric returns for the pooled sample as well as for specific groups. These are the most comparable to the parametric results in Table 1. However, given that our nonparametric estimates are specific to a given value of x, we present the variance of the returns as well in the third table. Figures 1-5 complement this table by displaying the density plots of the estimates for both the pooled sample and specific groups over time.

Similar to Table 2, the fourth table gives the cohort results for specific groups at the median values. Noting that the densities in Figures 1-5 cover the negative region, Table 5 gives the percentages of individuals which possess negative rates of return of education in each particular figure. The final table looks at the characteristics of individuals who possess negative returns and compares those to individuals who possess positive returns to attempt to figure out which characteristics of individuals are more likely to be associated with a negative return of education.

5.1 Pooled results

The first column of Table 3 displays the nonparametric estimates of the return of education across years. Each of the median estimates is large and positive. Unlike their parametric counterparts, nonparametric results do not show a clear increasing trend; the estimates fluctuate around 9%. On the other hand, the dispersion of the return of education is relatively constant over at roughly 0.02, although it is smaller in 1940 and 2005.

Figure 1 plots the densities of the nonparametric estimates. None of these densities appear to be normal and there appears to be a significant amount of skewness. Specifically, there are many individuals whose rates of return lie to the right of the mode(s). Another interesting observation is that there are a number of estimates which lie in below zero. These density plots imply that there exists a significant percentage of individuals having negative rates of return. The percentage of male workers who receive negative returns was only 3% in 1940, but rose dramatically to 9% in 1950 and to 16% in 1970. This number slightly decreased to 7% in 1970, but increased to 10% in 1980 and remained unchanged through 2000. 2005 saw a decrease of roughly 6%.

5.2 Sample splits

Although we are able to get a separate rate of return of education for each x and thus observe heterogeneity across the pooled sample, we are unaware of how the rate of return of education varies across and within specific groups. Again, we consider several age groups, two different groups based on race as well as splits by marital status and whether or not an individual was born in the United States. Specifically, we take the sample and split it according to one of the aforementioned categories. We then run separate nonparametric regressions for each group for each year and report the median estimate in the first panel of Table 2 and the variance of the estimates across the group in the second panel. A more comprehensive description of the results can be found in Figures 2-5 where we give the density plots for each group in each time period.

5.2.1 Age groups

In Table 3, columns 2-5 give the median rates of return of education by age group. In general, we find that younger workers have higher rates of return than older workers, although in 1970, 1990, and 2000, we do find that individuals aged 30-39 have the highest rate of return. The variance of the return of education is consistently smaller across years for individuals aged 30-39 than that for other age groups.

Figure 2 displays the density plot of these estimates. The densities show that there is a larger fraction of older workers who have negative return of education. In particular, except in 1940 when the rates of return were extremely high, the percentage of individuals with negative returns for individuals aged 40 and above were consistently around 20% across time and even reached 34% in 1980 when there was a sharp decrease in rates of return. While there were less individuals aged 30 and below receiving negative returns, the percentage was still high, ranging from 4% to 19%. In contrast, there is only a small fraction of individuals aged 30-40 that had negative returns; during the period 1980 - 2000, the percentage was either near zero or around one-percent. We also see in the figure that the variance is smaller for workers between the ages of 30 and 40. This is likely partially due to the fact that these workers are more homogenous in their ages as opposed to the other two groups.

5.2.2 Race

The next two columns of the table give the results by race. First, we again find that black men have higher rates of return than do white men. Moreover, comparing parametric and nonparametric results, we notice that for blacks, parametric estimates are generally smaller than their nonparametric counterparts, indicating that the parametric estimates are downward biased. Second, in most years, the variance of the return of education is smaller for blacks than for whites.

The density plots of these estimates are displayed in Figure 3. The results indicate that except in 1940, a smaller fraction of black men have negative rates of return as compared to white men. In 1960 and 1970, the fraction of individuals having negative returns increased for both white and black men, but the increase was larger for white men. We also see that the density is generally shifted to the right for black men and hence corresponds to the higher rate of return at the median value. At the same

16

time, other than in 1980 and 2005, there appears to be little difference in terms of the variance of the rates of return between the two groups.

5.2.3 Marital status

Columns 8 and 9 of Table 3 present the results by marital status. First, comparing the parametric and nonparametric results, we notice that for married males, parametric estimates are generally larger than their nonparametric counterparts. In contrast, for non-married males, parametric estimates are considerably smaller than their non-parametric counterparts. Such discrepancies indicate that the returns to education are underestimated (overestimated) for non-married (married) individuals. The large differences in the magnitudes of the estimates also imply that the gap in the returns to education between married and non-married males are also underestimated over time using the parametric method. Second, while the variance of the returns to education remains relatively constant over time for married males, there exists substantial variation in the variance for non-married males.

In terms of the density plots (Figure 4), there are several findings of note. First, in contrast to the results by age groups and those by race, the results by marital status indicate that the percentages of individuals with negative return of education are actually similar between married and non-married groups, although there appears to be some differences in 1990 and 2005. Second, the density plots of nonparametric estimates for married males appear to be unimodal across years, while those for the non-married show moderate evidence of bimodality during the period 1960-1980. A possible explanation for this bimodality can come from our definition of not married. This includes workers who have never been married, those who have been divorced and those who are widowers. Each of these groups perhaps should be treated separately in subsequent analysis.

5.2.4 Immigration status

The final two columns of Table 3 present the results by immigration status. First, the nonparametric results indicate that the median return of education is lower for immigrants than for natives in most years, consistent with the parametric results. However, we notice that for immigrants, parametric estimates are smaller than their nonpara-

17

metric counterparts. Such a discrepancy indicates that the return of education is consistently underestimated for immigrants in the parametric model. Nonparametric results also show a different picture of the trend over time. Except in 1940, the return of education is relatively stagnant for whites, whereas there is a steady increase for immigrants from 1950 to 1980; eventually the return of education for immigrants exceeded that for natives by 0.8%. However, such a trend did not persist. After 1980, the gap between immigrants and natives appeared again; the gaps were about 4% in 1990 and 2005. Also, not only were the estimates for immigrants smaller than those for natives, the variance of the returns to education are also smaller than those for natives.

In terms of the density plots (Figure 5), two points stand out. First, there has been a declining trend of the percentage of individuals with negative rates of return, for both natives and immigrants, although we see a slight increase in 1970 for natives and in 1980 for immigrants. Second, while the percentages of negative returns for immigrants were larger than those for natives during the period 1940-1980, such a difference was reversed after 1980. However, we note that at the same time, the median rate of return was higher for natives. Taking a look at the second panel of Table 3 shows that the variance of the returns to immigrants dropped dramatically after 1980 and thus likely explains this result. However, the question remains as to why this variance fell.

5.3   Cohort results

As we argued in the previous section, comparison between age groups in a particular year may not necessarily reflect the life cycle pattern for the returns to education. Here we follow the same cohort over time to get a better understanding of the life cycle pattern. The results for cohorts by group are presented in Table 4.

5.3.1   Race

The results by race are presented in the first two panels of Table 4. The nonparametric results for cohorts are consistent with their parametric counterparts: individuals have larger rates of return when they are younger, although there appears to be some exceptions in 1980. The results hold for both white and black men. However, the

nonparametric results for the cross-sectional analysis are less clear; there are several exceptions where the largest return is observed for older groups. For instance, in 1940, black men aged 40-49 had a return of 19.2% whereas individuals aged 20-29 had 19% and those aged 30-39 had 17%. One possible explanation for this strange result is the relatively small sample size for that particular time period.

5.3.2 Marital status

The results by marital status are presented in the third and fourth panels of Table 4. We again find a declining return of education over the life cycle. Although the magnitude of the values differ, these trends are more or less in line with those in the parametric regressions. The one major departure from the parametric results are that married individuals who are aged 50-59 have a higher rate of return than their non-married counterparts. In the parametric models, the higher rate of return for the non-married is consistently higher. Even more dramatic is that there are two cases (1970 and 2000) where the median rate of return of education is actually negative. We will discuss the presence of negative returns more carefully in the next section.

5.3.3 Immigration status

The results by immigration status are presented in the final two panels of Table 4. We again find a declining return of education over the life cycle. Again, the results indicate that the decline in the return of education is smaller for immigrants than for natives. However, we find even more cases where the median rate of return is negative for individuals who are aged 50-59. We show three cases for immigrants (1960, 1970 and 1990) and one case for natives (1970).

5.4 Negative returns

Perhaps the most novel results of this paper are with respect to the discovery of negative rates of return of education. To the best of our knowledge, we believe that no one has discovered negative rates of return of education. That being said, an occasional rate of return of education with nonparametric methods should not be alarming or out of the ordinary given the small sample biases of nonparametric estimators. However, if we find a large proportion of the estimates to be negative,

19

this would suggest that these negative rates do exist. In this sub-section we try to figure out how many individuals have negative rates of return of education and then we attempt to uncover who they are and what characteristics they have in common.

5.4.1 What percentage are negative?

Table 5 shows the percentage of observations which have negative rates of return for a particular group in a particular year. These approximately represent the area to the left of zero in Figures 1-5. It is obvious from first glance at the table that the percentages vary greatly across the time periods and groups. For instance, for individuals aged 30-39, the percentage is roughly five percent or lower. This perhaps can be ignored. That being said, many of these returns are near zero and this is an interesting finding in itself. At the same time, individuals in excess of age 40 comprise large amounts of negative rates of return. Further examination into this group shows that the percentages are highest for the oldest members of this group.

Other percentages are noteworthy. Except in 1940, the percentage of negative rates of return are higher for whites than blacks. Except in 2000, the percentage of negative rates of return are higher for married individuals than those who are not married. Finally, until 1990, a higher percentage of negative rates of return were possessed by immigrants. At that point, the trend switched to natives being more likely to be negative. This is worth further investigation as trends in immigration from the country of origin may have an impact on the rate of return to schooling. We discuss this in more detail later.

5.4.2 Who do the negatives belong to?

The lengthy Table 7 tries to uncover whether particular groups have more negative or positive rates of return of education with respect to other attributes. Specifically, each value in the table represents the difference in the percentage of individuals who have negative return from those who have positive returns for a specific attribute. For example, the first element in the table says that in 1940, for people aged 20-29, the difference between the percent married who have a negative rate and the percent married who have a positive return is 11.7%. In other words, for those people aged 20-29, you are more likely to have a negative rate of return of education if you are

20

married. Further, given that this value is bold, it indicates that this difference is significant at (at least) the 10% level.

We consistently find that older individuals are more likely to have negative rates of return (positive values in Table 7). This result is expected given the results for the cohorts.

Except for individuals aged 30-39 (before 2000) and those who are not married, having a child under the age of five makes it more likely to have a negative rate of return. Possible explanations for this phenomenon include fewer hours worked and a decrease in productivity that occurs when more time is devoted to young children (e.g., lack of sleep). It is, however, interesting to note that starting in 2000, individuals aged 30-39 are more likely to have a negative return when they have children under the age of 5. This perhaps reflects the trend in the delay of marriage and thus the delay in having children.

Another interesting finding which is consistent with the percentages discussed above is that immigrants are more likely to possess negative returns for all but a single group (black). This result appears to be strongest for individuals aged 30-39. This perhaps reflects that education obtained in a foreign country may not carry over international borders, whereas younger individuals are more likely to have more education in the United States. Foreign education is also likely for older individuals, but combining this with the fact that they are older and hence are twice as likely to have negative returns to education perhaps makes the reported difference smaller.

Finally, we had an interest in the region of origin for the immigrants and how this affected negative returns. We considered several regions, but report results for the Americas, Europe and Oceania, and Asia. Starting in 1980, if the immigrant came from the America's then they were more likely to have a negative rate of return. This difference was generally significant. There seems to be no discernible pattern for Europe and Oceania, but it is consistently found that being from an Asian country makes it less likely to have negative rates of returns, but this result is only significant slightly over half of the time.[8]

---

[8]We also examine characteristics such as geographical region and self-employed, but few results were significant and few to no consistent patterns existed.

6   Where do we go from here?

As motivated in the introduction, we have simply given a set of descriptive statistics. The results here tell us what the rates of return are for particular groups over time. We have found that the rates of return differ significantly over time, across groups and within groups. However, we have often given citations to where these results are in line with past studies, but none of the results here tell us why. The point of this exercise was to showcase heterogeneity in the rate of return of education and to attempt to uncover who had a higher rate of return. During this exercise we also uncovered that a significant proportion of the rates of return were negative. We further attempted to figure out which groups were most represented in this infamous group.

Further research should focus on allowing for the possibility of these heterogenous effects and uncovering why particular groups differ in terms of their rates of return of education. Also, we think more emphasis should be placed on heterogeneity within groups as opposed to across groups. It has been argued in this paper that recessions caused a drop in the rate of return of education for that 1980. This suggests that labor demand related variables as well as region/state effects may also be useful in uncovering the true rates of return of education. Finally, research should also be conducted to see how these results differ in other developed as well as lesser developed and emerging countries.

7   Conclusion

In this paper we examined heterogeneity in rates of return of education both across and within groups by employing a nonparametric approach. Specifically, we used generalized kernel estimation to examine the rate of return of education from 1940-2005 using U.S. Census data. Our results uncovered significant heterogeneity both across and within groups. Many of the results found (on average) were in line with previous studies. For instance, we found the rate of return of education to be larger for blacks than for whites. However, we also found inconsistencies. In that same example, we found that the difference in the rate of return between blacks and whites (on average) was much larger than previously thought. Finally, we uncovered that

a significant portion of the population had negative rates of returns to education. We attempted to uncover which groups were more likely to possess this newfound phenomenon.

Overall, we cannot place much emphasis on any particular point estimate from this study. The rates of return are likely impacted by omitted variables, measurement error and the lack of repeated observations for specific individuals. Although our results are somewhat raw, they do provide a springboard for further studies on the estimation of rates of return of education. We suggest that when estimating Mincer type regressions that authors focus on both heterogeneity across and within groups. Further, we suggest that future studies also focus on macroeconomic variables which are likely to have an affect on the perceived rate of return of education.

References

[1] Angrist J, Chen S. 2007. "Long-Term Effects of Vietnam-Era Conscription: Schooling, Experience and Earnings," Center for Economic Studies, U.S. Census Bureau Working Paper.

[2] Card D, Krueger A. 1992. "School Quality and Black-White Relative Wages: A Direct Assessment," *Quarterly Journal of Economics* 107: 151-200.

[3] Card D, Lemieux T. 2000. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis," NBER Working Paper No. 7655.

[4] Chiswick B. 1978. "The Effects of Americanization on the Earnings of the Foreign-born," *Journal of Political Economy* 86: 897–922.

[5] Chiswick B, Miller P. 2004. "Language Skills and Earnings among Legalized Aliens," in K. Zimmermann and A. Constant (eds.), *How Labor Migrants Fare.* Springer-Verlag: Berlin.

[6] Clark RL, Mitchell OS. 2005. *Reinventing the Retirement Paradigm.* Oxford University Press: Oxford.

[7] Fan J, Gijbels I. 1992. Variable Bandwidth and Local Linear Regression Smoothers. *Annals of Statistics* 20: 2008-2036.

[8] Goldin C, Katz LF. 1999. "The Returns to Skill in the United States across the Twentieth Century," NBER Working Papers No. 7126.

[9] Harmon C, Hogan V, Walker I. 2003. "Dispersion in the Economic Return to Schooling," *Labour Economics* 10: 205-214.

[10] Heckman J, Lochner L, Todd P. 2008. "Earnigns Functions and Rates of Return," *Journal of Human Capital* 2: 1-31.

[11] Heckman J, Polachek S. 1974. Empirical Evidence on Functional Form of the Earnings-Schooling Relationship," *Journal of the American Statistical Association* 69: 350-354.

[12] Hsiao C, Li Q, Racine J. 2007. A Consistent Model Specification Test with Mixed Discrete and Continuous Data. *Journal of Econometrics* 140: 802-826.

[13] Hurvich CM, Simonoff JS, Tsai CL. 1998. Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society, Series B* 60*: 271-293.

[14] Li Q, Racine J. 2004. Cross-Validated Local Linear Nonparametric Regression. *Statistica Sinica* 14: 485-512.

[15] Li Q, Racine J. 2007. *Nonparametric Econometrics: Theory and Practice.* Princeton University Press: Princeton.

[16] Mincer J. 1974. *Schooling, Experience, and Earnings.* Columbia University Press: New York.

[17] Pagan A, Ullah A. 1999. *Nonparametric Econometrics.* Cambridge University Press: Cambridge.

[18] Racine J, Li Q. 2004. Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data. *Journal of Econometrics* 119: 99-130.

[19] Rivera-Batiz F. 2007. "How Do Migrants from Latin America and the Caribbean Fare in the US Labour Market?" *World Economy* 30: 1399-1429.

[20] Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall: London.

[21] Soldo BJ, Mitchell OS, Tfaily R, McCabe JF. 2006. "Cross-Cohort Differences in Health on the Verge of Retirement" NBER Working Paper No. 12762.

[22] Welch F. 1973. "Black and White Differences in Returns to Schooling," *American Economic Review* 63: 893-907.

Data Appendix

Coding scheme for years of schooling in 1990 - 2005

Category	Mid-point
No schooling completed	0
Nursery school	0
Kindergarten	0
1st to 4th grade	2.5
5th grade to 8th grade	6.5
9th grade	9
10th grade	10
11th grade	11
12th grade, no diploma	12
High school graduate	12
Some college	13
Associate degree	14
Bachelor's degree	16
Master's degree	18
Professional degree	20
Doctorate degree	20

Table 1: Parametric results over time

Year	All	20-29	30-39	40-49	50-59	White	Black	Married	Not Married	Native	Immigrant
1940	0.132	0.187	0.143	0.107	0.081	0.122	0.102	0.141	0.12	0.149	0.057
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
1950	0.09	0.106	0.092	0.079	0.065	0.079	0.089	0.102	0.073	0.1	0.031
	(0.001)	(0.003)	(0.002)	(0.003)	(0.004)	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)
1960	0.103	0.131	0.095	0.088	0.081	0.094	0.1	0.143	0.079	0.11	0.046
	(0)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0)	(0.001)	(0.001)
1970	0.104	0.125	0.092	0.085	0.075	0.101	0.102	0.159	0.073	0.109	0.067
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0)	(0.001)	(0.002)
1980	0.095	0.098	0.093	0.077	0.059	0.092	0.109	0.142	0.063	0.098	0.074
	(0)	(0.001)	(0.001)	(0.001)	(0.001)	(0)	(0.001)	(0)	(0)	(0)	(0.001)
1990	0.114	0.164	0.113	0.101	0.07	0.113	0.135	0.159	0.081	0.123	0.086
	(0)	(0.001)	(0.001)	(0.001)	(0.001)	(0)	(0.001)	(0.001)	(0)	(0)	(0)
2000	0.113	0.164	0.12	0.095	0.074	0.113	0.143	0.152	0.083	0.125	0.09
	(0)	(0.001)	(0.001)	(0.001)	(0.001)	(0)	(0.001)	(0.001)	(0)	(0)	(0)
2005	0.121	0.189	0.13	0.113	0.082	0.121	0.159	0.152	0.09	0.141	0.088
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)

Table 2: Parametric results by cohort

				Black				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.131	0.113	0.151	0.147	0.153	0.214	0.211	0.255
Age 30-40	0.091	0.077	0.093	0.088	0.104	0.138	0.141	0.155
Age 40-49	0.083	0.066	0.075	0.08	0.081	0.112	0.112	0.128
Age 50-59	0.075	0.065	0.072	0.08	0.058	0.07	0.088	0.091
				White				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.177	0.095	0.116	0.115	0.09	0.159	0.156	0.179
Age 30-40	0.132	0.082	0.085	0.086	0.088	0.108	0.119	0.129
Age 40-49	0.095	0.07	0.078	0.079	0.073	0.097	0.092	0.114
Age 50-59	0.071	0.05	0.068	0.067	0.054	0.062	0.066	0.077

				Married				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.172	0.088	0.102	0.094	0.08	0.13	0.131	0.148
Age 30-40	0.139	0.086	0.088	0.088	0.086	0.101	0.112	0.12
Age 40-49	0.106	0.078	0.083	0.082	0.074	0.095	0.088	0.107
Age 50-59	0.085	0.06	0.077	0.071	0.055	0.064	0.067	0.074
				Not Married				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.168	0.092	0.136	0.139	0.091	0.157	0.16	0.176
Age 30-40	0.128	0.084	0.112	0.103	0.101	0.118	0.119	0.125
Age 40-49	0.093	0.067	0.103	0.094	0.084	0.111	0.102	0.112
Age 50-59	0.058	0.062	0.082	0.09	0.066	0.072	0.073	0.086

				Immigrant				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.137	0.102	0.085	0.108	0.094	0.119	0.138	0.13
Age 30-40	0.079	0.081	0.075	0.074	0.085	0.092	0.101	0.103
Age 40-49	0.052	0.057	0.069	0.069	0.068	0.088	0.084	0.086
Age 50-59	0.037	0.023	0.053	0.059	0.057	0.072	0.074	0.071
				Native				
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.184	0.106	0.131	0.122	0.098	0.172	0.169	0.208
Age 30-40	0.149	0.09	0.096	0.094	0.093	0.117	0.129	0.143
Age 40-49	0.128	0.082	0.089	0.087	0.078	0.104	0.099	0.125
Age 50-59	0.113	0.079	0.084	0.077	0.058	0.067	0.069	0.084

Table 3: Median and the variance of the nonparametric rate of return estimates over time

Median estimate

Year	All	< 30	30-40	> 40	White	Black	Married	Not Married	Native	Immigrant
1940	0.144	0.198	0.123	0.113	0.142	0.189	0.119	0.206	0.165	0.075
1950	0.091	0.111	0.081	0.082	0.077	0.113	0.081	0.164	0.091	0.043
1960	0.083	0.137	0.063	0.063	0.079	0.112	0.073	0.189	0.1	0.058
1970	0.1	0.117	0.084	0.088	0.098	0.115	0.086	0.303	0.099	0.07
1980	0.095	0.148	0.1	0.05	0.105	0.114	0.092	0.146	0.093	0.101
1990	0.131	0.132	0.135	0.096	0.134	0.208	0.127	0.172	0.143	0.103
2000	0.098	0.124	0.154	0.071	0.093	0.151	0.099	0.114	0.129	0.114
2005	0.151	0.167	0.134	0.119	0.133	0.214	0.122	0.197	0.15	0.106

Variance of estimates

Year	All	< 30	30-40	> 40	White	Black	Married	Not Married	Native	Immigrant
1940	0.008	0.006	0.002	0.015	0.011	0.038	0.006	0.009	0.011	0.005
1950	0.016	0.027	0.003	0.032	0.007	0.005	0.015	0.038	0.019	0.021
1960	0.023	0.015	0.004	0.021	0.025	0.023	0.014	0.037	0.024	0.005
1970	0.02	0.023	0.001	0.028	0.027	0.031	0.019	0.049	0.025	0.01
1980	0.02	0.013	0.001	0.046	0.032	0.008	0.014	0.052	0.026	0.02
1990	0.019	0.011	0.002	0.02	0.018	0.019	0.013	0.014	0.015	0.002
2000	0.02	0.02	0.002	0.018	0.013	0.008	0.014	0.032	0.017	0.002
2005	0.012	0.019	0.015	0.029	0.025	0.009	0.018	0.011	0.02	0.003

Table 4: Nonparametric Results by Cohort

	White							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.178	0.092	0.117	0.115	0.13	0.141	0.147	0.133
Age 30-40	0.135	0.075	0.068	0.086	0.079	0.138	0.129	0.174
Age 40-49	0.113	0.079	0.048	0.056	0.066	0.12	0.088	0.129
Age 50-59	0.106	0.058	0.043	0.031	0.039	0.086	0.074	0.103
	Black							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.19	0.154	0.153	0.196	0.139	0.277	0.191	0.257
Age 30-40	0.17	0.11	0.109	0.098	0.112	0.176	0.155	0.26
Age 40-49	0.192	0.093	0.068	0.088	0.072	0.128	0.123	0.218
Age 50-59	0.178	0.082	0.03	-0.064	0.036	0.154	0.056	0.15

	Married							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.148	0.083	0.122	0.084	0.12	0.164	0.15	0.156
Age 30-40	0.122	0.085	0.068	0.081	0.107	0.142	0.103	0.132
Age 40-49	0.097	0.079	0.059	0.065	0.103	0.128	0.085	0.118
Age 50-59	0.116	0.071	0.005	0.057	0.041	0.076	0.108	0.109
	Not Married							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.21	0.188	0.197	0.307	0.135	0.187	0.108	0.236
Age 30-40	0.191	0.094	0.102	0.205	0.061	0.136	0.186	0.186
Age 40-49	0.179	0.08	0.023	0.095	0.073	0.128	0.07	0.146
Age 50-59	0.103	0.058	0.009	-0.04	0.034	0.113	-0.034	0.088

	Immigrant							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.118	0.081	0.108	0.128	0.137	0.126	0.15	0.16
Age 30-40	0.09	0.056	0.07	0.088	0.108	0.108	0.119	0.134
Age 40-49	0.07	0.037	0.048	0.062	0.09	0.064	0.109	0.097
Age 50-59	0.071	0.012	-0.044	-0.039	0.078	-0.04	0.114	0.088
	Native							
Age Group	1940	1950	1960	1970	1980	1990	2000	2005
Age 20-30	0.186	0.115	0.133	0.121	0.126	0.152	0.187	0.176
Age 30-40	0.139	0.072	0.105	0.08	0.073	0.138	0.159	0.193
Age 40-49	0.15	0.084	0.092	0.081	0.099	0.019	0.101	0.151
Age 50-59	0.129	0.015	0.022	-0.047	0.064	0.05	0.08	0.105

Table 5: Percentage of Negative Returns by Years

Year	All	< 30	30-40	> 40	White	Black	Married	Non-Married	Native	Immigrant
1940	0.03	0.02	0.01	0.12	0.05	0.23	0.04	0.03	0.05	0.12
1950	0.09	0.14	0.05	0.25	0.05	0.01	0.14	0.15	0.13	0.23
1960	0.16	0.08	0.05	0.23	0.19	0.12	0.14	0.11	0.10	0.20
1970	0.07	0.12	0.04	0.21	0.14	0.14	0.18	0.15	0.14	0.14
1980	0.10	0.14	0.00	0.34	0.22	0.05	0.16	0.17	0.13	0.16
1990	0.10	0.13	0.01	0.20	0.10	0.04	0.09	0.03	0.06	0.05
2000	0.10	0.19	0.00	0.18	0.05	0.03	0.11	0.12	0.07	0.02
2005	0.04	0.04	0.05	0.17	0.10	0.06	0.13	0.02	0.07	0.02

Table 6: SUMMARY STATISTICS

Year	Variable	Age 20-30	Age 31-40	Age 41-50	Age 51-60	White	Black	Married	Not Married	Immigrant	Native
1940	Married	**0.17**	-0.071	-0.006	0.019	**0.117**	-0.02	**-0.153**	**-0.004**	0.011	**0.116**
	Child ¡ 5	**0.135**	0.061	-0.018	-0.004	**-0.103**	**0.096**	0.03	0.001	**-0.057**	**-0.083**
	Black	**0.154**	**0.181**	-0.001	0.023	**0.15**	-0.007	**0.208**	**0.187**	-0.006	**0.076**
	Immigrant	**0.083**	**0.322**	**0.089**	0.017	**-0.261**	0.173	-0.178	-0.521	-0.082	**-0.223**
	Age 20-30					-0.078	-0.048	-0.143	-0.146	-0.2	-0.165
	Age 31-40					-0.086	-0.065	-0.114	0.126	0.019	0.129
	Age 41-50					0	0	0	0	0	0
	Age 51-60										
	America									-0.034	
	Europe & Oceania									0.035	
	Asia									0.008	
1950	Married	**-0.078**	-0.008	-0.017	-0.007	**0.032**	0.164	**-0.106**	**-0.005**	-0.021	**0.06**
	Child ¡ 5	**-0.121**	**0.123**	-0.01	0.001	**-0.07**	**-0.195**	0.008	**0.042**	0.016	**-0.088**
	Black	0.01	0.032	**-0.028**	-0.009	0	0	0.008	**0.081**	-0.005	0
	Immigrant	-0.006	0.014	0.005	-0.007	**0.066**	0.056	**0.042**	**0.081**	0	0
	Age 20-30					**-0.078**	**-0.27**	**-0.065**	**-0.13**	**0.104**	**-0.173**
	Age 31-40					**-0.086**	**-0.252**	**-0.124**	-0.011	-0.021	**-0.023**
	Age 41-50					**-0.14**	**-0.215**	**-0.083**	**0.175**	**-0.148**	**-0.051**
	Age 51-60									**0.013**	**-0.01**
	America									**0.05**	
	Europe & Oceania									**-0.051**	
	Asia									0	
1960	Married	**-0.09**	-0.031	-0.007	**-0.025**	**0.032**	0.033	**-0.142**	**-0.006**	-0.021	0.007
	Child ¡ 5	**-0.109**	0.035	**-0.029**	**-0.013**	**-0.067**	**-0.195**	0.007	0.011	**-0.167**	**-0.119**
	Black	0.009	-0.01	0.019	0.003	**0.047**	0.021	**0.053**	**0.092**	-0.009	0
	Immigrant	0.014	**0.046**	0.032	**0.023**	**-0.083**	**-0.244**	**-0.065**	**-0.3**	**-0.17**	-0.101
	Age 20-30					-0.04	**-0.191**	**-0.188**	-0.007	**-0.211**	-0.079
	Age 31-40					**-0.072**	0.076	**-0.045**	-0.008	**-0.187**	-0.093
	Age 41-50					-0.011	-0.011	-0.005	-0.04	-0.005	0.006
	Age 51-60									**-0.097**	
	America									**0.22**	
	Europe & Oceania									**-0.055**	
	Asia									0	
1970	Married	**-0.04**	-0.018	-0.003	-0.013	**0.064**	0.037	**-0.081**		0.028	**0.095**
	Child ¡ 5	**-0.048**	**0.177**	0.016	-0.001	**-0.044**	**-0.141**	-0.001	0	**-0.22**	**-0.046**
	Black	-0.018	-0.027	-0.015	**0.039**			-0.001	0.018	-0.011	-0.014

Table 6 – (cont.) Summary Statistics

Year	Variable	Age 20-30	Age 31-40	Age 41-50	Age 51-60	White	Black	Married	Not Married	Immigrant	Native
	Immigrant	0.009	0.064	-0.017	0.017	0.021	-0.001	0.009	0.009		-0.181
	Age 20-30					-0.12	-0.294	-0.043	-0.139	-0.17	0.071
	Age 31-40					0.066	-0.084	-0.031	0.103	-0.244	-0.076
	Age 41-50					-0.013	0.2	-0.097	0.059	-0.199	0.011
	Age 51-60					-0.004	-0.029	0.005	-0.032	-0.01	
	America									-0.075	
	Europe & Oceania									0.229	
	Asia									-0.082	
1980	Married	-0.155	0.196	0.053	0.001	0.058	0.201			0.049	0.134
	Child ¡ 5	-0.102	0.073	0.005	0.004	-0.102	-0.139	-0.188	0	-0.078	-0.082
	Black	-0.004	0.112	0.018	0.005			-0.01	-0.024	-0.039	-0.007
	Immigrant	0.015	0.561	0.049	-0.004	0.003	-0.058	0.012	0.015		
	Age 20-30					-0.096	-0.379	-0.21	-0.051	-0.23	-0.225
	Age 31-40					-0.101	-0.271	-0.176	0.065	0.05	-0.079
	Age 41-50					0.05	-0.111	0.041	0.076	0.012	0.022
	Age 51-60					0	0	0	0	0	0
	America									0.109	
	Europe & Oceania									-0.067	
	Asia									-0.02	
1990	Married	-0.26	-0.13	-0.027	0.026	0.082	0.183			0.271	0.24
	Child ¡ 5	-0.16	0.169	0.044	-0.01	-0.116	-0.164	-0.202	0.006	-0.204	-0.154
	Black	-0.002	-0.062	-0.024	-0.008			-0.015	-0.011	-0.028	-0.021
	Immigrant	0.034	0.485	0.125	0.001	0.034	-0.043	0.069	0.13		
	Age 20-30					-0.137	-0.308	-0.156	-0.401	-0.3	-0.268
	Age 31-40					-0.244	-0.291	-0.251	-0.131	-0.308	-0.265
	Age 41-50					-0.027	-0.15	-0.097	-0.008	-0.215	-0.198
	Age 51-60					0	0	0	0	-0.036	-0.023
	America									0.051	
	Europe & Oceania									0.127	
	Asia									-0.159	
2000	Married	-0.206	-0.144	0.023	-0.019	0.162	0.276			0.257	0.193
	Child ¡ 5	-0.104	-0.195	0.045	-0.001	-0.131	-0.164	-0.14	-0.01	-0.04	-0.118
	Black	-0.009	-0.072	-0.016	0.016			-0.016	0.003	-0.064	-0.024
	Immigrant	0.033	0.789	0.045	0.025	0.049	-0.077	0.058	0.03		
	Age 20-30					-0.145	-0.272	-0.03	-0.24	-0.248	-0.201
	Age 31-40					-0.236	-0.306	-0.248	-0.16	-0.08	-0.207

32

Table 6 – (cont.) Summary Statistics

| Year	Variable	Age 20-30	Age 31-40	Age 41-50	Age 51-60	White	Black	Married	Not Married	Immigrant	Native
	Age 41-50					-0.216	-0.236	-0.097	0.042	-0.24	-0.214
	Age 51-60					0	0	0	0	0	0
	America									0.184	
	Europe & Oceania									-0.014	
	Asia									-0.183	
2005	Married	0.225	0.036	0.016	0.038	0.159	0.212			0.033	0.18
	Child ¡5	0.182	-0.044	0.002	0.001	-0.059	-0.129	-0.111	0.007	-0.169	-0.088
	Black	-0.055	-0.039	-0.029	-0.01		-0.038	-0.016	0.025	-0.042	-0.019
	Immigrant	0.396	0.416	0.195	0.02	0.03		0.038	0.114		
	Age 20-30					-0.123	-0.221	-0.033	-0.337	-0.13	-0.176
	Age 31-40					-0.108	-0.245	-0.165	-0.042	-0.294	-0.097
	Age 41-50					-0.15	-0.268	-0.243	-0.022	-0.243	-0.217
	Age 51-60					0	0	0	0	0	0
	America									0.373	
	Europe & Oceania									-0.104	
	Asia									-0.232	

33
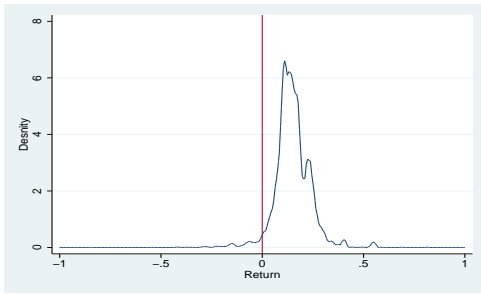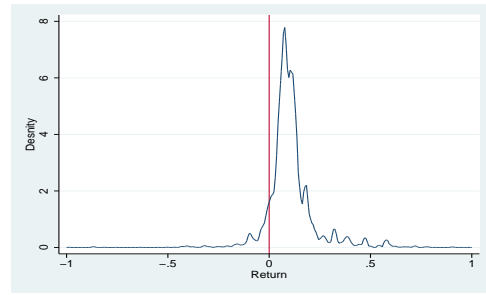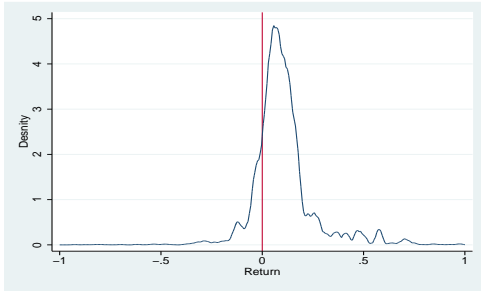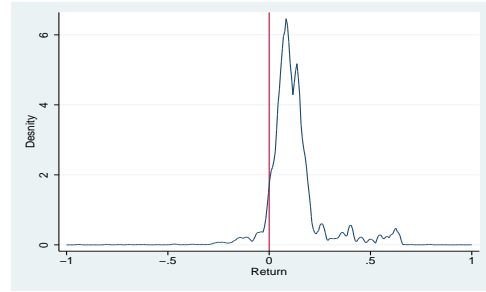
(a) 1940
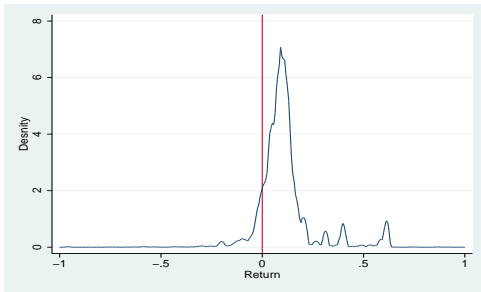
(b) 1950
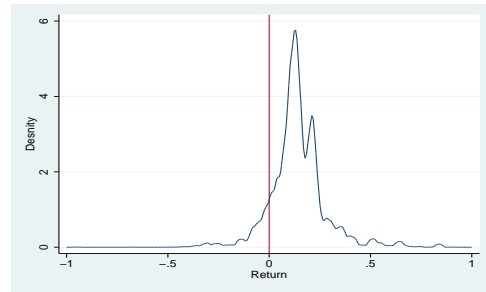
(c) 1960
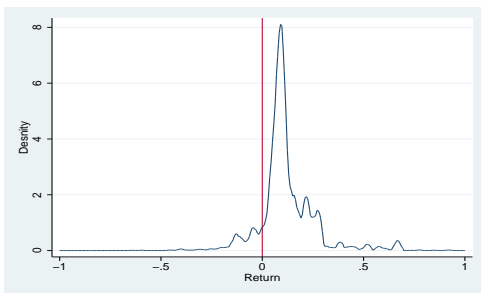
(d) 1970
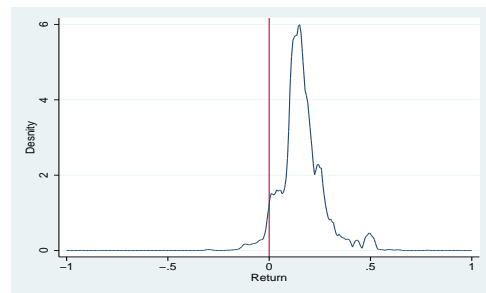
(e) 1980

(f) 1990

(g) 2000

(h) 2005

Figure 1: Nonparametric Results by Year
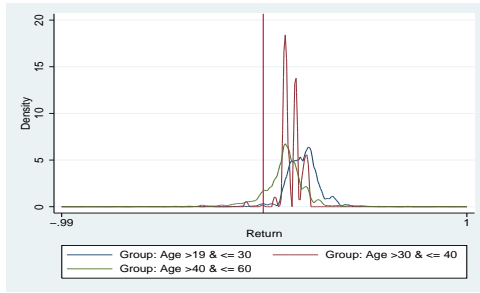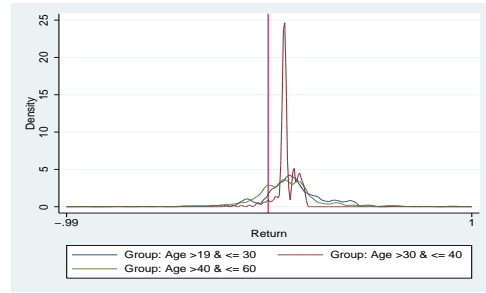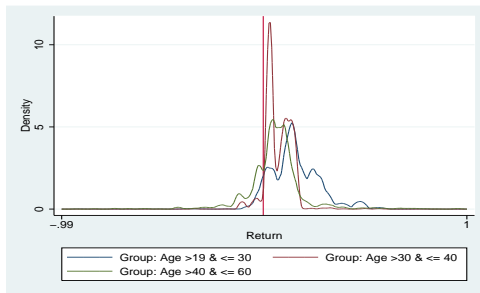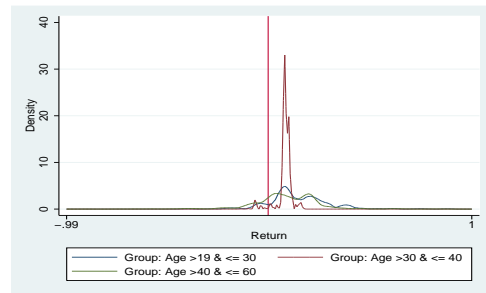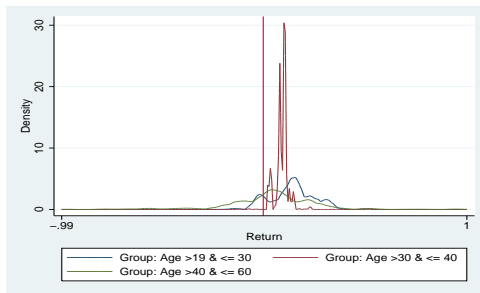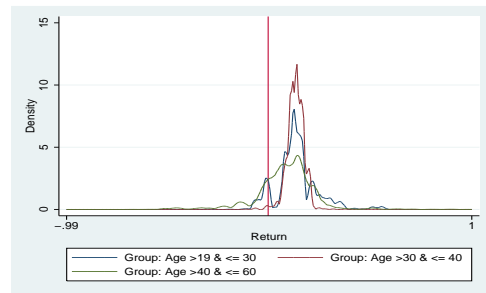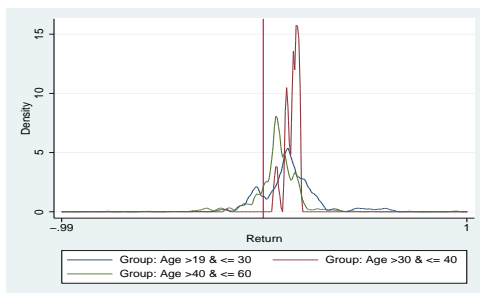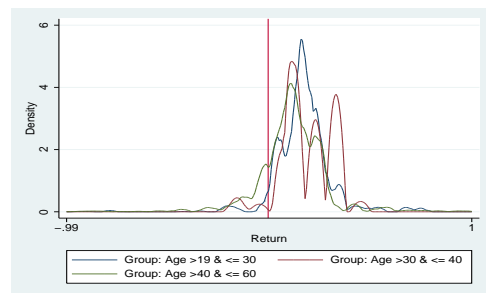
Figure 2: Nonparametric Results by Year and Age Groups
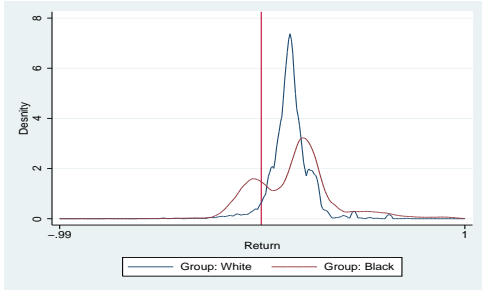
(a) 1940
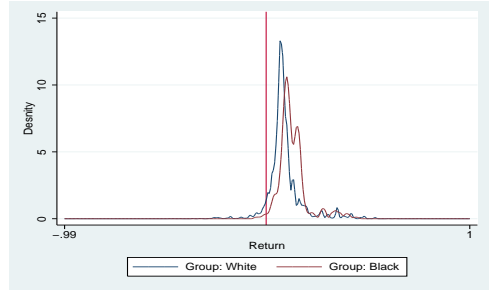
(b) 1950
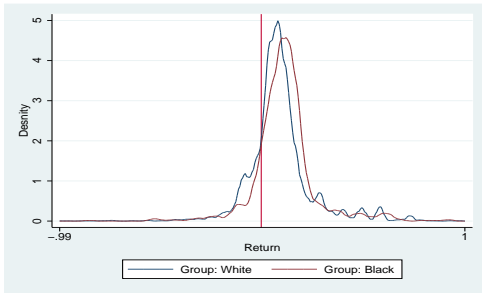
(c) 1960
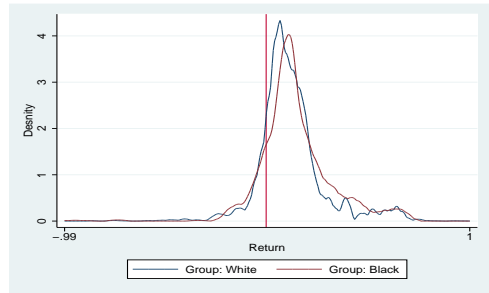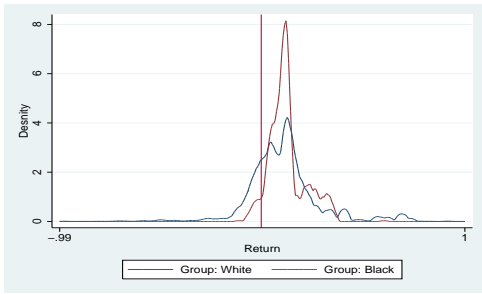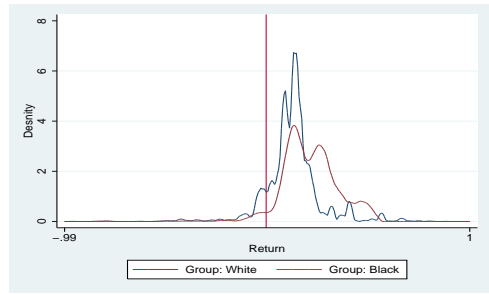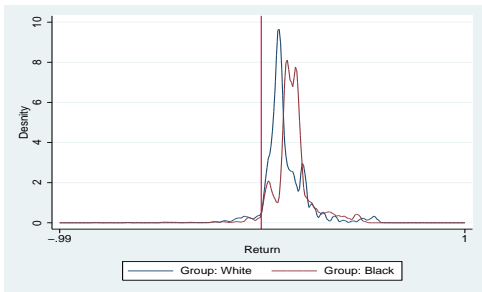
(d) 1970
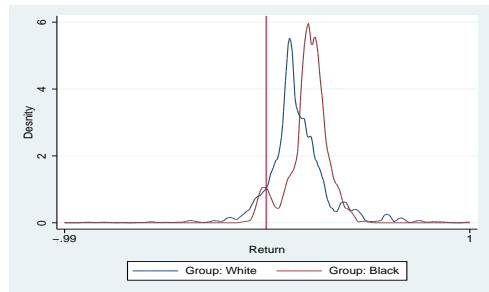
(e) 1980

(f) 1990

(g) 2000

(h) 2005

Figure 3: Nonparametric Results by Year and Race
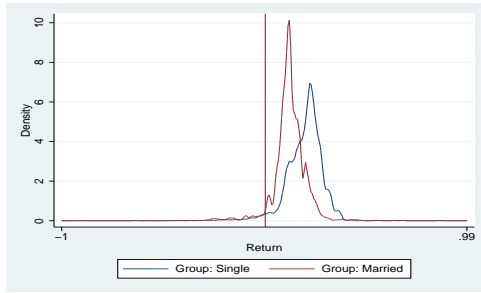
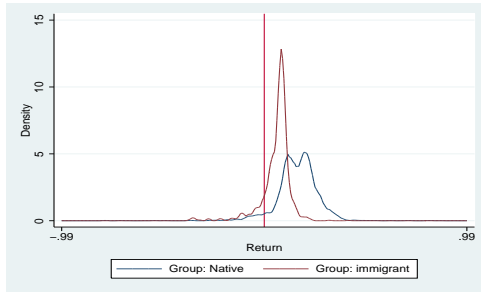(a) 1940
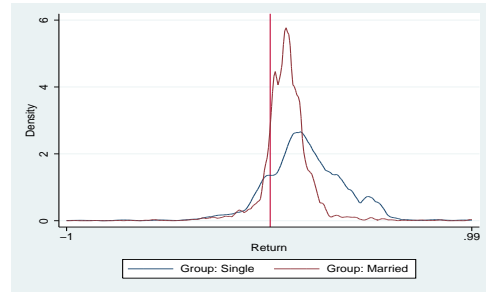
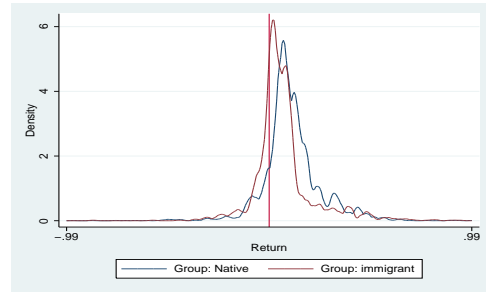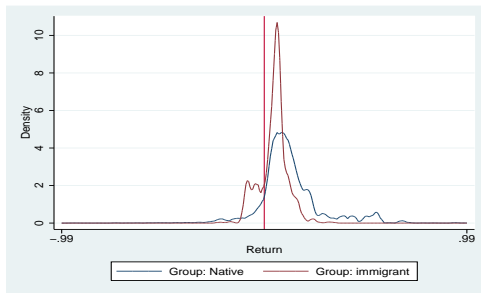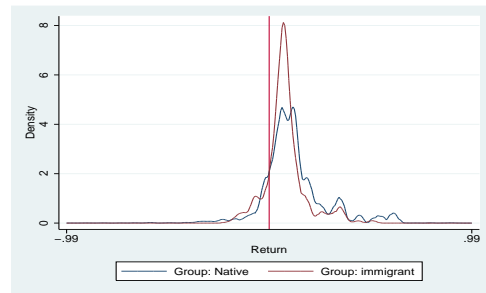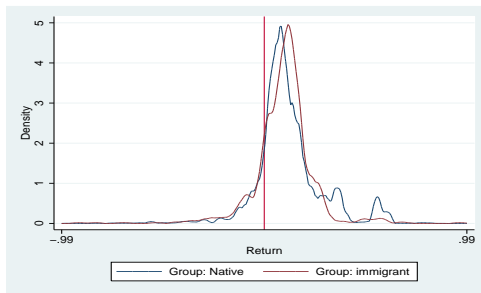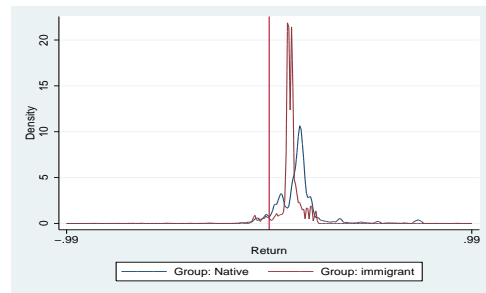(b) 1950
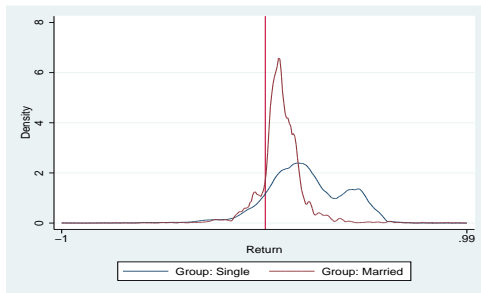
(c) 1960
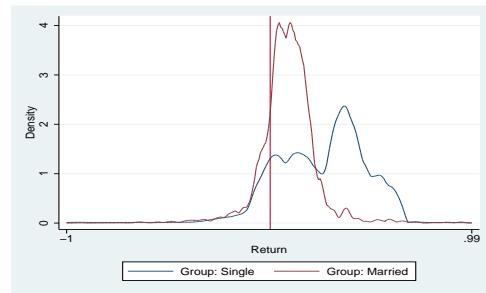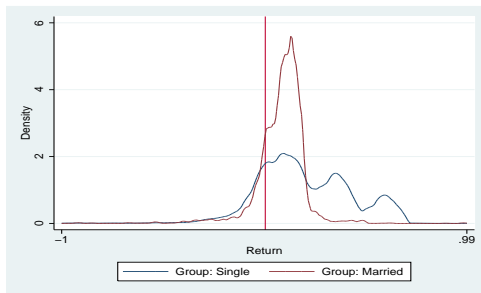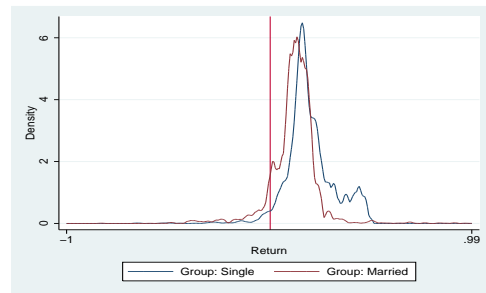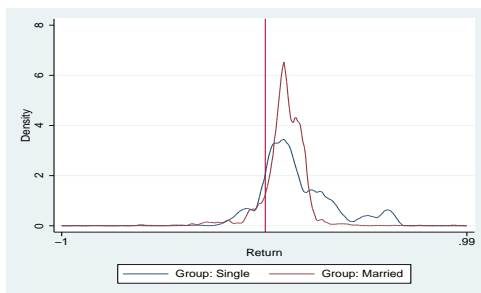
(d) 1970
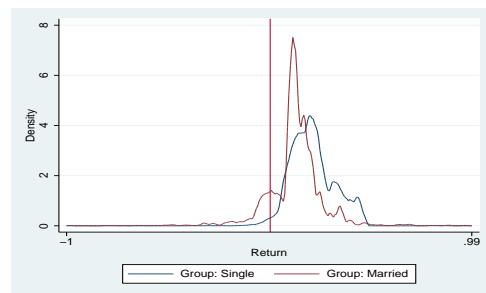
(e) 1980

(f) 1990

(g) 2000

(h) 2005

Figure 4: Nonparametric Results by Year and Marital Status
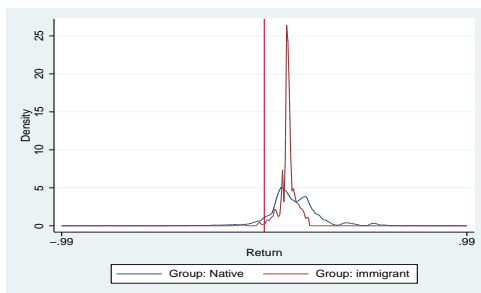
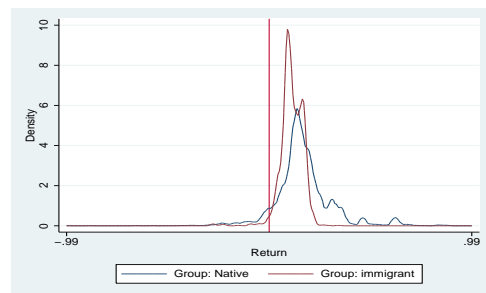37

(a) 1940

(b) 1950

(c) 1960

(d) 1970

(e) 1980

(f) 1990

(g) 2000

(h) 2005

Figure 5: Nonparametric Results by Year and Immigration Status