THE DISTRIBUTION OF RETURNS TO EDUCATION

by

ANNE-CHARLOTTE SOUTO

DANIEL J. HENDERSON, COMMITTEE CHAIR
PAUL PECORINO
CARY DECK
LE WANG
FIRAT SOYLU

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Economics, Finance, and Legal Studies
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2019

Copyright Anne-Charlotte Souto 2019
ALL RIGHTS RESERVED

ABSTRACT

In this work, we revisit the traditional human capital framework and infer that risk as measured by the shape of the returns to education's distributions should be included. While education is often considered to be an investment good, human capital models often ignore the impact of risk on education investment decisions. This thesis has two aims. First, we want to find out how our different measures of risk evolved through time and between different groups. Second, we want to find out if those risks impacted education investment decisions through changes in the expected returns. That is, we investigate whether there exist risk-return trade-offs in education.

In the first chapter, we overview nonparametric (spline and kernel) regression methods and illustrate how they may be used in labor economic applications. We focus our attention on issues commonly found in the labor literature such as how to account for endogeneity via instrumental variables in a nonparametric setting. We showcase these methods via data from the Current Population Survey.

In the second chapter, we estimate the risk-return trade-off in the context of education. If education is treated like any other investment good, risk could play an important role in individuals educational decisions. As portfolio theory predicts, there could be a trade-off between returns to education and risks concerning the returns: higher risks are generally associated with higher returns. We contribute to the literature by proposing various measures of risk based on the distribution of returns to education, which are in turn based on nonparametric regression results using the Current Population Survey dataset (1980-2015).

We infer that risk-averse individuals prefer distributions with positive skewness and low kurtosis. Our results confirm the findings of the literature, i.e. we observe compensation for variance. We also find statistically significant compensation for the higher moments: skewness and kurtosis. Interestingly, we find that the relationship between expected returns and the higher moments skewness and kurtosis is non-linear.

In the third chapter, we build on the second chapter to test two hypothesis: first whether there is heterogeneity in the risk of educational investments and if so whether there is compensation for that risk. We use our individual-level estimated rates of return to education and split them in three different ways: by occupation, by region and race, and by region and education-level. We infer that there is heterogeneity, not only in the expected returns (1st moment), but also in the risk faced by individuals (higher moments). We also add to the second chapter by testing whether risk-return trade-offs exist between occupations, whites and non-whites, and different education-level. We expect, for example, occupations that retain higher risk to be compensated by higher mean returns.

Generally, we find risk-return trade-offs exist between states, occupations, whites and non-whites, and different education-level, for all three measures of risk. Surprisingly, we find that kurtosis matters more than skewness as a measure of risk. Moreover, the trade-offs between skewness, kurtosis, and expected returns are not always in the directions predicted by theory on decision making under uncertainty.

DEDICATION

To my mother and father, who have always supported me and encouraged me to continue in my education. A special thanks to my mother who spent countless hours listening during the times of stress. Thank you for your encouragement through all of these years of graduate school and for being a pillar of support in all aspects of my life. To my step-father, for his tremendous help in all things great and small.

ACKNOWLEDGMENTS

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of my doctorate. I would like to express my sincere gratitude to my advisor Dr. Daniel J. Henderson for his continuous support of my Ph.D study and related research, and his dedication to my professional development. I am grateful for your mentorship as a graduate student and I hope to continue working with you for years to come.

I am also thankful for the help of the other members of my committee, Dr. Paul Pecorino, Dr. Le Wang, Dr. Cary Deck, and Dr. Firat Soylu, and the many other University of Alabama faculty members and staff for all of their guidance and assistance over the past several years. I am grateful to all those who gave me comments on my research and job market materials. I would like to extend a special thanks to Dr. Paul Pecorino for his enthusiasm and support as a research mentor.

I would like to thank my fellow doctoral students for their feedback, cooperation and, of course, friendship. In addition, I would like to express my gratitude to Shamar Stewart, Tina Zhang, and Kaveh Hasani, for their emotional support and their continuous willingness to help with assignments, coding, and writing.

Last but not the least, I would like to thank my family: my parents, my step-father, and my brother and sisters for supporting me throughout writing this thesis and in my life in general. A special thanks to my step-father Pascal Orliac for always being willing to proofread everything I write and listen to my presentations, and for his countless useful comments, remarks, and suggestions.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
AN INTRODUCTION TO NONPARAMETRIC REGRESSION FOR LABOR ECONOMISTS (WITH DANIEL J. HENDERSON)

1.1 Introduction

This survey aims to (re-)introduce applied labor economists to nonparametric regression techniques. Specifically, we discuss both spline and kernel regression, in an approachable manner. We present an intuitive discussion of estimation and model selection for said methods. We also address the use of nonparametric methods in the presence of endogeneity, a common issue in the labor literature, but seldom accounted for in applied nonparametric work.

Accounting for endogeneity is well understood in the parametric literature once a suitable instrument is obtained. Standard methods have been around for some time, but these methods do not always transfer in a straightforward manner in the nonparametric setting. This has caused many to shy away from their use, even with the knowledge that this can lead to additional insight (Henderson and Parmeter (2015)).

To showcase these methods, we will look at the relationship between experience, education and earnings. We will begin by ignoring the endogeneity of education and then will discuss how to control for this via a nonparametric control function approach. While nonparametric estimation may seem like a choice, it should be stated that the parametric alternative requires strict functional form assumptions, which if false, likely lead to biased and inconsistent estimators. In practice, the functional relationship between education and earnings as well as between education and its instruments is typically unknown. By using nonparametric regression, we relax these functional form restrictions and are more likely to uncover the causal relationship.

To empirically illustrate these methods, we use individual-level data obtained from the March Current Population Survey (CPS) to highlight each concept discussed. To eliminate additional complications, we primarily focus on a relatively homogeneous sub-group, specifically, working age (20 to 59 years old) males with four-year college degrees.

In what follows, we first slowly introduce the fundamentals of spline and kernel estimators and then discuss how to decide upon various options of each estimator. This should build the foundation for understanding the more advanced topic of handling endogenous regressors. By illustrating these techniques in the context of labor-specific examples, we hope that this helps lead to widespread use of these methods in labor applications.

1.2 Nonparametric Regression

In a parametric regression model, we assume which functional form best describes the relationship between the response and explanatory variables. If this form is correct, and the remaining Gauss-Markov assumptions hold, we will have unbiased and efficient estimators. However, if these assumptions do not hold, these estimators are likely biased and inconsistent. Nonlinear parametric models exist, but are often complicated to estimate and still require *a priori* knowledge of the underlying functional form.

Nonparametric regression offers an alternative. The methods discussed here estimate the unknown conditional mean by using a "local" approach. Specifically, the estimators use the data near the point of interest to estimate the function at that point and then use these local estimates to construct the global function. This can be a major advantage over parametric estimators which use all data points to build their estimates (global estimators). In other words, nonparametric estimators can focus on local peculiarities inherent in a data set. Those observations which are more similar to the point of interest carry more weight in the estimation procedure.

This section will introduce two commonly used nonparametric techniques, and will provide the notation and concepts that will be used for the remainder of this review. Specifically, we discuss spline and kernel regression estimation. To help bridge gaps, we make connections to well-known techniques such as ordinary and weighted least-squares.

1.2.1 Spline Regression

Spline regression can be thought of as an extension of ordinary least-squares (OLS). Consider the basic univariate linear model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \ i = 1, 2, \ldots, n, \tag{1.1}$$

where for a sample of n observations, y is our response variable, x is our explanatory variable, ϵ is our usual error term and we have two parameters: a constant and a slope (α and β, respectively). The right-hand side of (1.1) can be thought of as a linear combination of 1 and x, we call them the "bases" of the model. One popular way to transform (1.1) into a nonlinear function is to add higher-order polynomials. A quadratic model would add one extra basis function x^2 to the model, which corresponds to adding the term $\beta_2 x_i^2$ to (1.1). In matrix form, the number of bases would correspond to the number of columns in the matrix X:

$$y = X\beta + \epsilon, \tag{1.2}$$

where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

for the linear case (2 bases), and

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$$

for the quadratic case (3 bases).

These two cases are illustrated in Figure 1.1 where x is years of experience and y is the log wage (adjusted for inflation). To highlight a relatively homogeneous group, we restrict our sample to college-educated (16 years of schooling) males working in personal care and service (or related occupations) between 2006 and 2016.[1] For each panel, the solid line represents white males and the dashed line non-whites. Our linear model (i.e., OLS) shows a strong wage gap between whites and non-whites which seems to remain constant (in percentage terms) as the sale workers gain experience (i.e., similar slopes). Adding experience squared to the model (quadratic model) allows us to better capture the well-known nonlinear relationship between log wage and experience. As workers gain experience, we expect their log wage to increase, but at a decreasing rate. The quadratic model (bottom-left panel) shows a large increase in log wages early in a career with a slight downfall towards the end. Also, this model tends to suggest that the wage gap between white and non-white males working in personal care and service varies with experience. Non-white workers appear to have a more constant and slower increase in their predicted log wages.

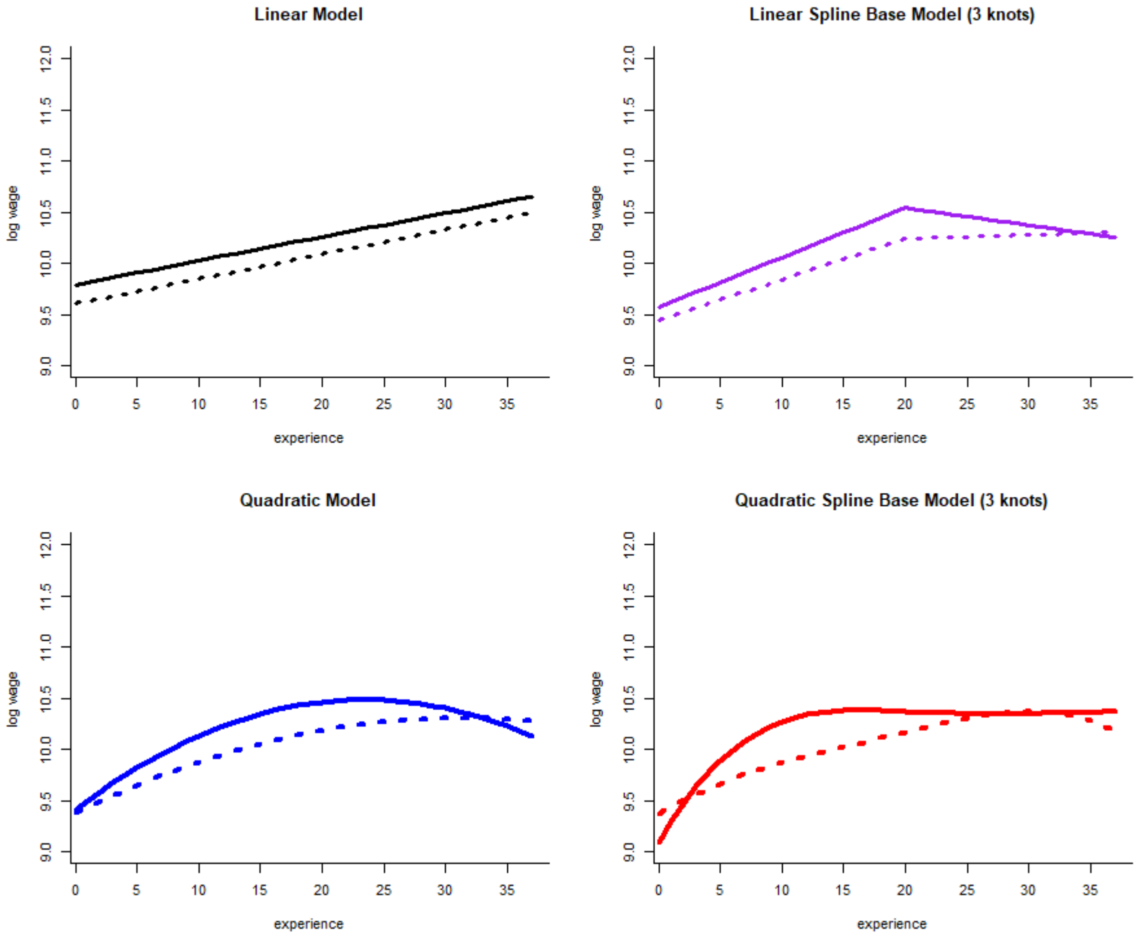
Linear Spline Bases

In our example, we could argue that although wages should increase with experience (increase in competence/knowledge), there may be a point where more experience will not increase wages or perhaps even decrease it (slower cognitive ability/decreases in efficiency). Suppose we created a model with the equivalent of two linear regression:

---

[1] Fixing the sample to college educated males allows us to plot these figures in two dimensions.

Figure 1.1: Log-wages versus Experience for White versus Non-white College-educated Males Working in Personal Care and Service



one for the first 20 years of experience, and another for the latter years. This would be equivalent of adding the following basis function to our linear model:

$$(x - 20)_+,$$

where the + sign indicates that the function is set to zero for all values of x where $(x - 20)$ is negative. This model is sometimes called the *broken stick* model because of its shape, but more generally is referred to as a linear spline base model with 3 knots. The 3 knots are at 0 (minimum value), 20, and 37 (maximum value) years of experience. Note that the maximum and minimum values of x will always be consider to be knots. For example, the linear model in equation (9) has two knots. Here we arbitrarily fixed the middle knot at 20 years of experience. We will discuss which knots to select and how many to select in Section 3.

The *broken stick* model with a break at $x = 20$ is written as

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - 20)_+ + \epsilon_i \tag{1.3}$$

and is illustrated in upper-right panel of Figure 1.1. We see a similar result to the quadratic model, that is, for white workers, we see a strong increase in wages in the first part of their career followed by a smaller decrease towards the end of their career. That being said, we arbitrarily fixed the middle knot at 20 years of experience. Without strong reasons to do so, it is premature to say anything about when the increase in the log wage stops and when the decrease begins. Noting the aforementioned issue, we also observe the wage gap widen at first with experience, but then converge at higher levels of experience.

Figure 1.2 illustrates how adding knots at different values can change the results. We present a model with 5 knots at $x = 0, 10, 20, 30, 37$, and a model with 20 knots (every 2 years) at $x = 0, 2, 4, ..., 34, 36, 37$. In matrix form, equation (1.2), the X matrix with 5 knots is given as

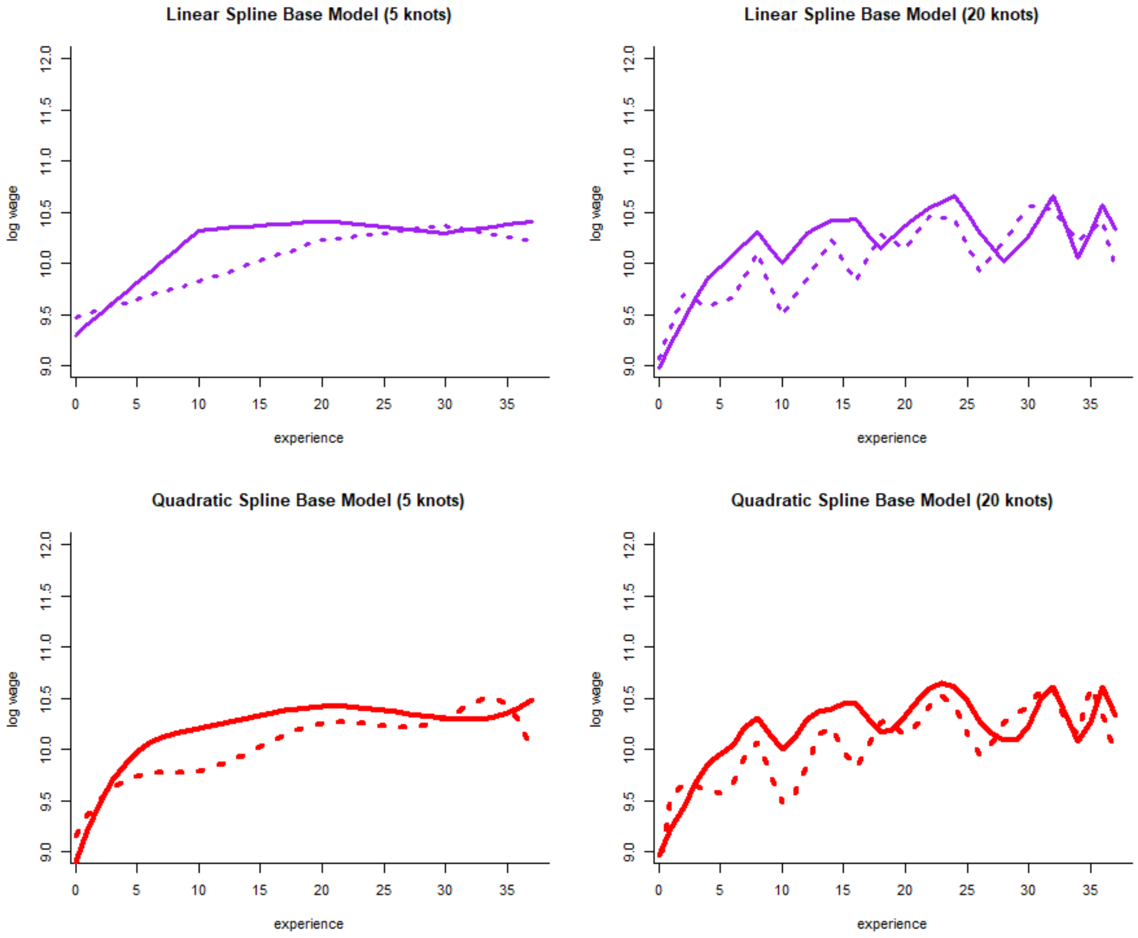
$$X = \begin{bmatrix} 1 & x_1 & (x_1 - 10)_+ & (x_1 - 20)_+ & (x_1 - 30)_+ & (x_1 - 37)_+ \\ 1 & x_2 & (x_2 - 10)_+ & (x_2 - 20)_+ & (x_1 - 30)_+ & (x_2 - 37)_+ \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & (x_n - 10)_+ & (x_n - 20)_+ & (x_1 - 30)_+ & (x_n - 37)_+ \end{bmatrix}$$

and with 20 knots,

$$X = \begin{bmatrix} 1 & x_1 & (x_1 - 2)_+ & ... & (x_1 - 36)_+ & (x_1 - 37)_+ \\ 1 & x_2 & (x_2 - 2)_+ & ... & (x_1 - 36)_+ & (x_2 - 37)_+ \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & (x_n - 2)_+ & ... & (x_1 - 36)_+ & (x_n - 37)_+ \end{bmatrix}$$

Adding knots at 10 and 30 years of experience allows the model to account for

Figure 1.2: Log-wage versus Experience for White versus Non-white College-educated Males Working in Personal Care and Service



the commonly seen mid-career flattening period. However, the function is still not very smooth and it is hard to tell from this model when log wages start to flatten out. Adding more knots allows for more flexibility, but this can potentially lead to overfitting. For example, in the linear base model with 20 knots (upper-right panel of Figure 1.2), the fitted line is appears to be modeling noise.

Quadratic Spline Bases

The linear spline base model is a combination of linear bases. The quadratic spline base model is a combination of quadratic bases. In other words, we simply add the corresponding squared function for each of the linear base functions. Consider our previous broken stick model with a middle knot at $x = 20$, we can transform it into a quadratic spline base model with a knot at $x = 20$ by replacing $(x-20)_+$ with the following bases:

$$x^2,\ (x-20)_+^2.$$

This quadratic spline base model is represented by the following equation

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 (x_i - 20)_+^2 + \epsilon_i, \tag{1.4}$$

and is illustrated in the bottom-right panel of Figure 1.1. We can see that the quadratic spline base model suggests a slightly different relationship between experience and log wage. The predicted log wage increases more dramatically for the first 5 years of work experience, but flattens out thereafter. The racial gap seems to be small at first, but widens greatly over the first 5 years. Non-white workers appear to slowly catch up over the course of their careers.

One of the main advantages of the quadratic over the linear spline base model is that it does not have any sharp corners (i.e., undefined gradients). It follows that for any number of knots, the resulting function will have continuous first derivatives. This is both a useful and aesthetically pleasing property. Adding more knots (lower-right panel of Figure 1.2) to the model adds more variability. It appears that for this example, 5 knots would be sufficient.

An important concept in economics (typically of secondary importance in statistics textbooks) is recovery of the gradients. In the linear case, the gradient is simply the estimated coefficient between two particular knots. In the quadratic (or higher-order) case, we use the same method to get the gradient as in a simple quadratic OLS model. The difference is that we calculate it between each knot. That is, to estimate a particular gradient for any type of spline model, we can simply take the partial derivative with respect to the regressor x. In its general form, our estimated gradient $\widehat{\beta}(x)$ for a particular regressor x is

$$\widehat{\beta}(x) = \frac{\partial \widehat{y}(x)}{\partial x}. \tag{1.5}$$

For our linear spline base example with 3 knots, this is

$$\widehat{\beta}(x) = \beta_1 + \begin{cases} \beta_2, & \text{if } x \in [20, 37) \\ 0, & \text{otherwise} \end{cases} \tag{1.6}$$

and for our quadratic spline base example with 3 knots

$$\widehat{\beta}(x) = \beta_1 + 2\beta_2 x_i + \begin{cases} 2\beta_3(x_i - 20), & \text{if } x \in [20, 37) \\ 0, & \text{otherwise} \end{cases} \tag{1.7}$$

B-Splines

We introduced linear and quadratic spline models with the truncated power basis function. Using the same truncated power functions, those models can be generalized to

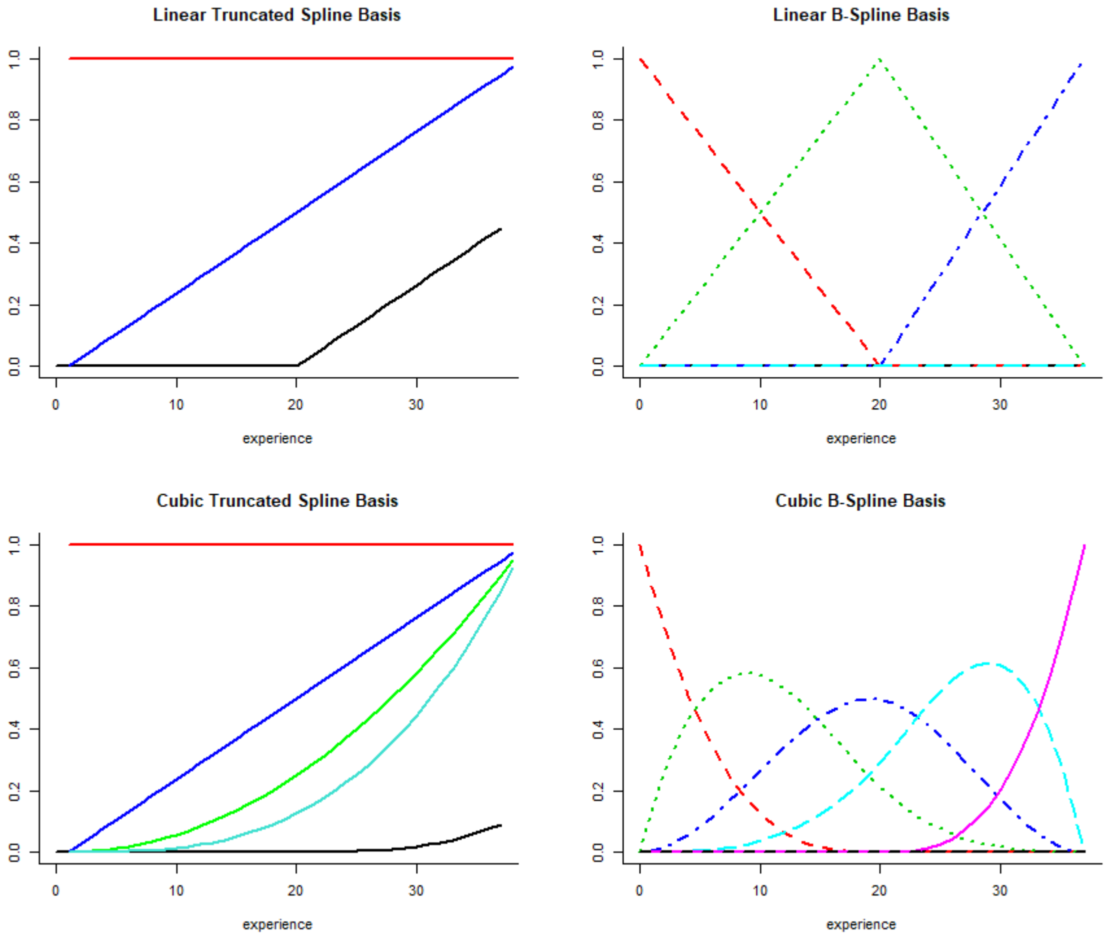
$$y = \beta_0 + \beta_1 x + \cdots + \beta_p x^p + \sum_{\mathbb{k}=1}^{\mathbb{K}} \beta_{p\mathbb{k}}(x - \kappa_{\mathbb{k}})_+^p + \epsilon_i, \tag{1.8}$$

where p is the degree of the power basis (truncated power basis of degree p). This generalizes our model by allowing for (1) other spline models (using p degrees), and (2) other bases for a given spline model (using $\mathbb{K}$ knots). This function has $p-1$ continuous derivatives and thus higher values of p should lead to "smoother" spline functions. Similar to before, the general form of the gradient is defined as

$$\widehat{\beta}(x) = \beta_1 + \cdots + p\beta_p x^{p-1} + \sum_{\mathbb{k}=1}^{\mathbb{K}} p\beta_{p\mathbb{k}}(x - \kappa_{\mathbb{k}})_+^{p-1}. \tag{1.9}$$

While this general form seems reasonable, splines computed from the truncated power bases in equation 1.8 may be numerically unstable. The values in the X-matrix may become very large (for large p), and the columns of the X-matrix may be highly correlated. This problem will only become worse with a higher number of knots. Therefore, (1.8) is rarely used in practice, but is instead typically transformed into equivalent bases with more stable numerical properties. One of the most popular is the *B-spline* basis.

Figure 1.3: Truncated and B-Spline Corresponding Bases with Knots at 0, 20, and 37 Years of Experience


Linear Truncated Spline Basis

Linear B-Spline Basis

Cubic Truncated Spline Basis

Cubic B-Spline Basis

This can be relatively difficult to present and code, but luckily there exist regression packages to easily transform the X-matrix into the more numerically stable version. Formally, we can compute the equivalence as

$$X_b = X_t L_p,$$

where X_t is a matrix of the bases (explanatory variables) used in (1.8) and L_p is a squared invertible matrix. The most commonly used transformation in the linear case is

$$B(x)_j = \begin{cases} \frac{x-\kappa_j}{\kappa_{j+1}-\kappa_j}, & \text{if } x \in [\kappa_j, \kappa_{j+1}) \\ \frac{\kappa_{j+2}-x}{\kappa_{j+2}-\kappa_{j+1}}, & \text{if } x \in [\kappa_{j+1}, \kappa_{j+2}) \\ 0, & \text{otherwise} \end{cases}$$

for $j = -1, 2, 3, ..., \mathbb{K}$.

To better illustrate this, consider our *broken stick* example from Figure 1.1: the linear-spline with one middle knot at 20 years of experience. The corresponding bases for this model are 1, x and, $(x - 20)_+$ and are shown in the upper-left panel of Figure 1.3. The B-spline transformation of the second knot (20 years of experience) for this example is

$$B(x)_{j=2} = \begin{cases} \frac{x-0}{20-0}, & \text{if } x \in [0, 20) \\ \frac{37-x}{37-20}, & \text{if } x \in [20, 37) \\ 0, & \text{otherwise} \end{cases}$$

The corresponding bases of this transformation are shown in the upper-right panel of Figure 1.3. $B(x)_{j=2}$ corresponds to the inverse V-shaped function which equals 1 when experience equals 20. The other two functions can be computed similarly using $j = -1$, and 3. Adding a higher degree to our model will change the shape of our basis functions. The two bottom panels of Figure 1.3 show the equivalent truncated spline basis and B-spline basis for the cubic case ($p = 3$).

While other basis functions exist (for example, radial basis functions), practitioners may prefer B-splines as they are both numerically more stable and relatively easy to compute. Both `R` and Stata packages are available. We used the now defunct *bs* $(\cdot)$ function in the *splines* package[2] in `R`. The *bspline* module is available in Stata for B-splines.

1.2.2 Kernel Regression

Instead of assuming that the relationship between y and x come from a polynomial family, we can state that the conditional mean is an unspecified smooth function

---

[2] See https://stat.ethz.ch/R-manual/R-devel/library/splines/html/bs.html and the seemingly equivalent *bSpline* $(\cdot)$ function in the *splines2* package.

$m(\cdot)$ and our model will be given as

$$y_i = m\left(x_i\right) + \epsilon_i,\ i = 1, 2, \ldots, n, \tag{1.10}$$

where the remaining variables are described as before. In much the same way spline regression can be thought of as an extension of OLS, kernel regression can be seen as an extension of WLS. That is, we are still minimizing a weighted residual sum of squares, but now we will weight observations by how close they are to the point of interest (i.e., a "local" sample). With spline regression, our local sample is defined as all the points included between two knots, where each point within that sample is weighted equally. Kernel regression goes a step further by estimating each point using a weighted local sample that is centered around the point of interest. The local sample is weighted using a kernel function, which possess several useful properties.

A kernel function defines a weight for each observation within a (typically) symmetric predetermined bandwidth. Unlike an OLS regression which makes no distinction of where the data are located when estimating the conditional expectation, kernel regression will estimate the point of interest using data within a bandwidth.

Before introducing the kernel estimators, let us first derive a kernel function. Consider x our point of interest; we can write an indicator function such that data fall within a range h (our bandwidth) around x:

$$n_x = \sum_{i=1}^{n} 1\left\{x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2}\right\},$$

The corresponding probability of falling in this box (centered on x) is thus n_x/n. This indicator function can be rewritten as

$$n_x = \sum_{i=1}^{n} \left(\frac{1}{2}\right) 1\left\{\left|\frac{x_i - x}{h}\right| \leq 1\right\}. \tag{1.11}$$

This function is better known as a uniform kernel and is more commonly written as

$$k(\psi) = \begin{cases} 1/2, & \text{if } |\psi| \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

where we have written $k(\psi)$ for convenience, where ψ is defined as $(x_i - x)/h$ and represents how "local" the observation x_i is relative to x. Though very simple and intuitive, the uniform kernel is not smooth. It is discontinuous at -1 and 1 (when the weight switches from $1/2$ to zero) and has a derivative of 0 everywhere except at theses two points (where it is undefined).

Table 1.1: Commonly used Second-order Kernel Functions

Kernel	$k(\psi)$	$\kappa_2(k)$
Uniform ($s = 0$)	$\frac{1}{2}\mathbf{1}\{\lvert\psi\rvert \leq 1\}$	1/3
Epanechnikov ($s = 1$)	$\frac{3}{4}(1-\psi^2)\mathbf{1}\{\lvert\psi\rvert \leq 1\}$	1/5
Biweight ($s = 2$)	$\frac{15}{16}(1-\psi^2)^2\mathbf{1}\{\lvert\psi\rvert \leq 1\}$	1/7
Triweight ($s = 3$)	$\frac{35}{32}(1-\psi^2)^3\mathbf{1}\{\lvert\psi\rvert \leq 1\}$	1/9
Gaussian ($s = \infty$)	$\frac{1}{\sqrt{2\pi}}e^{-(1/2)\psi^2}$	1

This kernel is rarely used, but it does possess some basic properties that we typically require of kernel functions. More formally, if we let the moments of the kernel be defined as

$$\kappa_j(k) = \int_{-\infty}^{\infty} \psi^j k(\psi) d\psi, \tag{1.12}$$

these properties are

1. $\kappa_0(k) = 1$ ($k(\psi)$ integrates to one),

2. $\kappa_1(k) = 0$ ($k(\psi)$ is symmetric), and

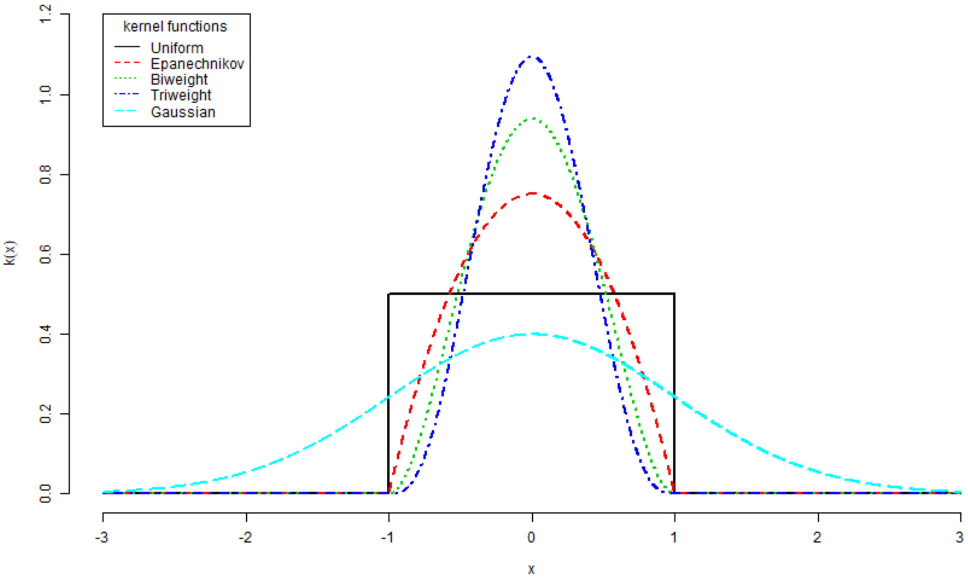
3. $\kappa_2(k) < \infty$ ($k(\psi)$ has a finite variance).

These are known as second-order kernels. In addition to the uniform kernel, several commonly known kernel functions can be found in Table 1.1 (with their second-moments) and Figure 1.4. Each of them are derived from the general polynomial family:

$$k_s(\psi) = \frac{(2s+1)!!}{2^{s+1}s!}(1-\psi^2)^s\mathbf{1}\{|\psi| \leq 1\}, \tag{1.13}$$

where !! is the double factorial. The most commonly used kernel function in econometrics is the Gaussian kernel as it has derivatives of all orders. The most commonly used kernel function in statistics is the Epanechnikov kernel function as it has many desirable properties with respect to mean squared error. We will discuss how to choose the kernel function and smoothing parameter (h) in Section 3.

Figure 1.4: Commonly used Second-order Kernel Functions



Local-Constant Least-Squares

The classic kernel regression estimator is the local-constant least-squares (LCLS) estimator (also known as the Nadaraya-Watson kernel regression estimator, see Nadaraya (1964) and Watson (1964)). While it has fallen out of fashion recently, it is useful as a teaching tool and still useful in many situations (e.g., binary left-hand-side variables).

To begin, recall how we construct the OLS estimator. Our objective function is

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - x_i\beta)^2,$$

which leads to the slope and intercept estimators, $\widehat{\beta}$ and $\widehat{\alpha}$.

Suppose instead of a linear function of x, we simply regress y on a constant (a). Our objective function becomes

$$\min_{a} \sum_{i=1}^{n} [y_i - a]^2,$$

which leads to the estimator $\widehat{a} = (1/n)\sum_{i=1}^{n} y_i = \bar{y}$. A weighted least-squares version of this objective function can be written as

$$\min_{a} \sum_{i=1}^{n} [y_i - a]^2 W(x_i),$$

where $W(x_i)$ is the weighting function, unique to the point x_i. If we replace the weighting function with a kernel function, minimizing this objective function yields the LCLS estimator

$$\widehat{a} = \widehat{m}(x) = \frac{\sum_{i=1}^{n} y_i k\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)}. \tag{1.14}$$

This estimator represents a local average. Essentially, we regress y locally, on a constant, weighting observations via their distance to x.

While equation 1.14 gives us the fit, economists are typically interested in the marginal effects (i.e., gradients). To estimate a particular gradient, we simply take the partial derivative of $\widehat{m}(x)$ with respect to the regressor of interest, x. Our estimated gradient $\widehat{\beta}(x)$ is thus

$$\widehat{\beta}(x) = \frac{\left(\sum_{i=1}^{n} y_i \frac{\partial k\left(\frac{x_i - x}{h}\right)}{\partial x}\right)\left(\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right) - \left(\sum_{i=1}^{n} y_i k\left(\frac{x_i - x}{h}\right)\right)\left(\sum_{i=1}^{n} \frac{\partial k\left(\frac{x_i - x}{h}\right)}{\partial x}\right)}{\left(\sum_{i=1}^{n} k\left(\frac{x_i - x}{h}\right)\right)^2}, \tag{1.15}$$

where, for example, $\frac{\partial k\left(\frac{x_i-x}{h}\right)}{\partial x} = \left(\frac{x_i-x}{h^2}\right) k\left(\frac{x_i-x}{h}\right)$ for the Gaussian kernel. Higher-order derivatives can be derived in a similar manner.

Local-Linear Least-Squares

While the LCLS estimator is intuitive, it suffers from biases near the boundary of the support of the data. As an alternative, most applied researchers use the local-linear least-squares (LLLS) estimator. The LLLS estimator locally fits a line as opposed to a constant.

The local-linear estimator is obtained by taking a first-order Taylor approximation of equation (1.10) via

$$y_i \approx m(x) + (x_i - x)\beta(x) + \epsilon_i,$$

where $\beta(x)$ is the gradient. Similar to the LCLS case, by labeling $m(x)$ and $\beta(x)$ as the parameters a and b, we get the following minimization problem

$$\min_{a,b} \sum_{i=1}^{n} [y_i - a - (x_i - x)b]^2 k\left(\frac{x_i - x}{h}\right),$$

which, in matrix notation (with q regressors) is

$$\min_{\delta} (y - X\delta)' K(x)(y - X\delta),$$

where $\delta = (a, b)'$, X is a $n \times (q+1)$ matrix with its ith row equal to $(1, (x_i - x))$ and $K(x)$ is a $n \times n$ diagonal matrix with its ith element equal to $k\left(\frac{x_i-x}{h}\right)$. This leads to the LLLS estimators of the conditional expectation ($\widehat{m}(x)$) and gradient ($\widehat{\beta}(x)$) as

$$\widehat{\delta}(x) = \begin{pmatrix} \widehat{m}(x) \\ \widehat{\beta}(x) \end{pmatrix} = (X'K(x)X)^{-1} X'K(x)y,$$

Notice that we can obtain the OLS estimator by replacing $K(x)$ by an identity matrix (giving all observations equal weight, i.e., bandwidth tending towards infinity), the weighted least-squares (WLS) estimator by replacing it with some other weighting function, and the generalized least-squares (GLS) estimator by replacing it with the inverse of the variance-covariance matrix of the errors (Ω).

Figure 1.5: Log-wage versus Experience for White versus Non-white College-educated Males Working in Personal Care and Service

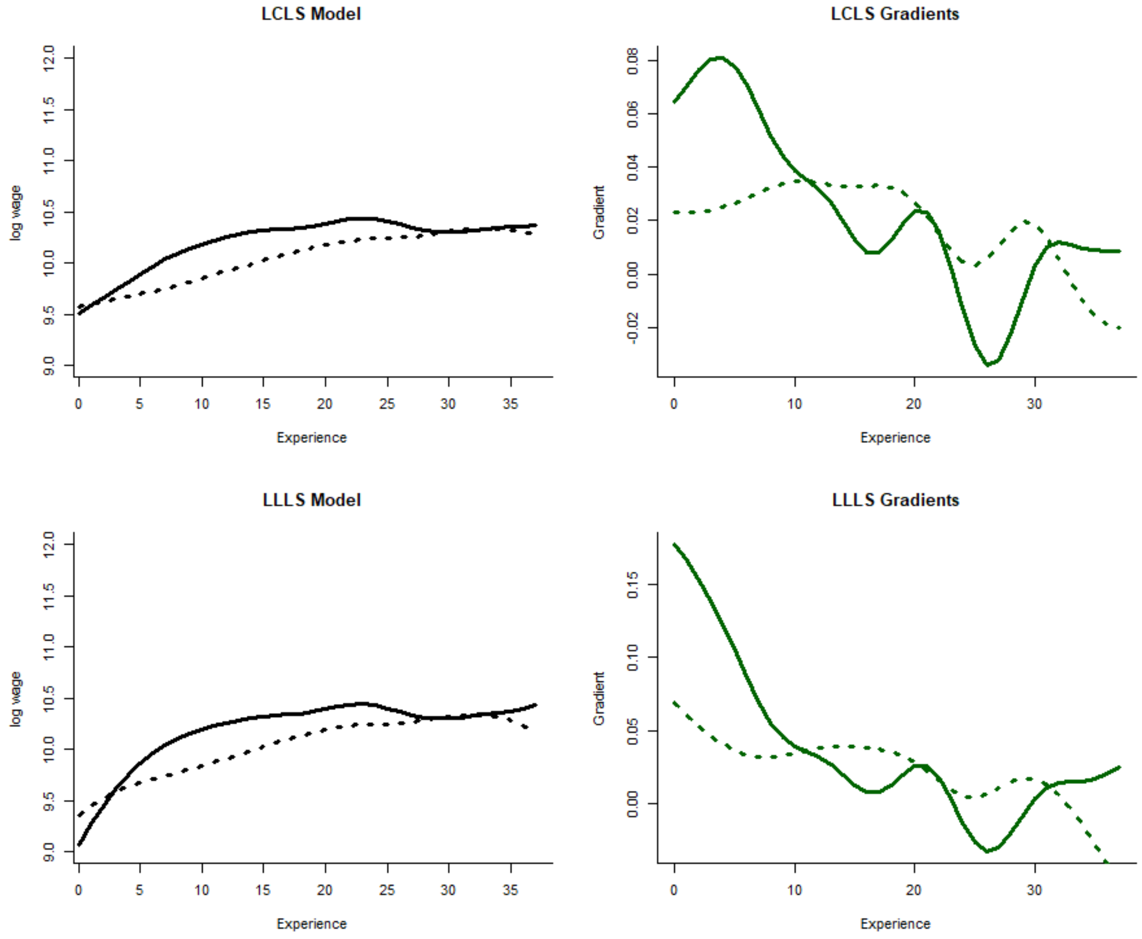


Figure 1.5 gives both the LCLS and LLLS estimates for white (solid line) and non-white (dashed line) college-educated males working in personal care and service. The gradients for each level of experience are also shown. Compared to the LCLS model, the LLLS model captures a stronger increase in log wage during the first 5 years of work experience with gradients ranging from 0.10 to 0.17. If taken literally, after only a year of working in personal care and service, white college-educated males wages increases by almost 17% on average while non-white college-educated males' wages increases by about 7%. The LCLS model, while showing a similar overall shape, shows

a much slower increase in those first few years of work experience with less than 4% increases in wages for non-whites and 5% to 8% increases for whites. Both models suggest that while white workers have much higher percent increases in their wages in the first few years, those year-to-year percent increases in their wages fall below non-white workers after 10 years of experience.

Local-Polynomial Least-Squares

The derivation of the LLLS estimator can be generalized to include higher-order expansions. The resulted family of estimators are called local-polynomial least-squares (LPLS) estimators. For the general case, if we are interested in the pth-order Taylor expansion, and we assume that the $(p+1)$th derivative of the conditional mean at the point x exists, we can write our equation as

$$y_i \approx m(x) + (x_i - x)\frac{\partial m(x)}{\partial x} + (x_i - x)^2 \frac{\partial^2 m(x)}{\partial x^2}\frac{1}{2!} + ... + (x_i - x)^p \frac{\partial^p m(x)}{\partial x^p}\frac{1}{p!} + \epsilon_i.$$

Replacing the parameters by $(a_0, ..., a_p)$, our kernel weighted least-squares problem can be written as

$$\min_{a_0,...,a_p} \sum_{i=1}^{n} [y_i - a_0 - (x_i - x)a_1 - (x_i - x)^2 a_2 - ... - (x_i - x)^p a_p]^2 k\left(\frac{x_i - x}{h}\right).$$

In matrix notation, our objective function becomes

$$\min_{\delta} (y - X\delta)' K(x)(y - X\delta)$$

where the only difference from the LLLS case ($p = 1$) is that the ith row of X is defined as $[1, (x_i - x), (x_i - x)^2, ..., (x_i - x)^p]$ and $\delta = (a_0, a_1, ..., a_p)'$. Minimizing the objective function leads to the local-polynomial least-square estimator

$$\widehat{\delta}(x) = \left(\widehat{m}(x), \frac{\partial \widehat{m}(x)}{\partial x}, \frac{\partial^2 \widehat{m}(x)}{\partial x^2}, ..., \frac{\partial^p \widehat{m}(x)}{\partial x^p}\right)' = (X'K(x)X)^{-1} X'K(x)y$$

The first question then becomes, how many expansions should we take? More expansions lead to less bias, but increased variability. This becomes a bigger problem when the number of covariates (q) is large and the sample size (n) is small. One promising data driven method to determine the number of expansions is considered in Hall and Racine (2015).

As is the case for splines, there exist options to employ these methods in popular software packages. In `R` we recommend the *np* package (Hayfield and Racine (2008)) and in Stata we recommend the *npregress* command.

1.3 Model Selection

For both spline and kernel regression, many seemingly arbitrary choices can greatly influence fit. The typical trade-off is between bias and variance. We want to make selections such that we avoid overfitting or underfitting. In this section, we first discuss penalty selection, knot selection, and degree selection in spline models; and then, kernel and bandwidth selection in kernel models.

1.3.1 Spline Penalty and Knot Selection

In Section 2.1, we saw that the fit is influenced by both our choice of degree of the piecewise polynomials, and by the number and locations of knots we include. However, in spline models, there is a third, more direct way, to influence fit: add an explicit penalty. In short, we want to select the degree of the piecewise polynomials, the knot locations, and the smoothing parameter λ (penalty) which best capture the underlying shape of our data. Though we will briefly discuss the selection of all three, it is easy to show that the choices of degree and knots are much less crucial than the choice of λ, the smoothing parameter (we will see a similar result for kernel regression). That is, when using a high enough number of knots and degrees, the "smoothness" of our fit can be controlled by λ. Hence, we will focus most of our discussion on the choice of λ when the degree and number of knots are fixed. Although there exist several ways to select our parameters in a data-driven manner, we will concentrate on one of the most commonly used approaches: cross-validation (CV).

Penalty Selection using Cross Validation

There are several ways to impose a penalty, but here we focus on a method that avoids extreme values (and hence too much variability). In a univariate setting using a linear spline, this penalty is

$$\sum \beta_{1\mathbb{k}}^2 < C,$$

where $\beta_{1\mathbb{k}}^2$ is the coefficient on the $\mathbb{k}$th knot[3]. In matrix form, our constrained objective function can thus be written as

$$\min_{\beta} \mid\mid y - X\beta \mid\mid^2 \text{ s.t. } \beta' D\beta \leq C,$$

and leads to Lagrangian[4]

$$\mathcal{L}(\beta, \lambda) = \min_{\beta,\lambda} \mid\mid y - X\beta \mid\mid^2 + \lambda^2 \beta' D\beta, \tag{1.16}$$

where D is a $(\mathbb{K}+2) \times (\mathbb{K}+2)$ matrix with diagonal $(0, 0, 1_1, ..., 1_{\mathbb{K}})$. Note that consistency will require that λ tends towards zero as the sample size (n) tends towards infinity.

The second term of (1.16) is called a roughness penalty because it penalizes through the value of the smoothing parameter (λ) the curvature of our estimated function. This type of regression is referred to as a penalized spline (p-spline) regression and yields the following solution and fitted values:

$$\widehat{\beta}_\lambda = (X'X + \lambda^2 D)^{-1} X' y$$

$$\widehat{y} = X(X'X + \lambda^2 D)^{-1} X' y.$$

To generalize these results to the pth degree spline model (equation (1.8)), we replace λ^2 by λ^{2p}[5] and transform the D-matrix: $D = \text{diag}(0_{p+1}, 1_{\mathbb{K}})$. A penalized B-

---

[3] The matrix of the coefficients being $\beta = [\beta_0, \beta_1, \beta_{11}, ..., \beta_{1\mathbb{K}}]'$

[4] Note that the last term $-\lambda^2 C$ disappears as it does not influence the solution

[5] We raise λ to the power of $2p$ because the way we add bases. Intuitively, raising λ to the power of $2p$ can be explained by the following example: if we transform X into αX for any $\alpha > 0$, we want to

spline (PB-spline) would simply include the transformation done to X (i.e., the square invertible matrix L_p) into the penalty term as well:

$$\widehat{y} = X_B(X_B'X_B + \lambda^{2p}L_p'DL_p')^{-1}X_B'y.$$

As $\lambda^{2p} \rightarrow \infty$ (infinite smoothing), the curvature penalty becomes predominant and the estimate converges to OLS. As $\lambda^{2p} \rightarrow 0$, the curvature penalty becomes insignificant. In this case, the function will become rougher (we will see a similar result with the bandwidth parameter for a LLLS regression). Figure 1.6 illustrates this effect using linear p-spline estimates for college-educated males working in personal care and service. The knots have been fixed at every five years of experience (0,5,10,...). As the penalty (λ) increases, it is clear that the fit becomes smoother and converges to an OLS estimate.

Figure 1.6 shows an intuitive fit of the data for a value of λ around 10. However, using a more systematic method to select λ would lead to less subjective and more comparable results. If we let $\widehat{m}(x_i;\lambda)$ be our nonparametric regression estimate at the point x with smoothing parameter λ, we can write a residual sum of squares objective function as

$$RSS(\lambda) = \sum_{i=1}^{n} \left[y_i - \widehat{m}(x_i;\lambda)\right]^2. \tag{1.17}$$

The problem with this approach, is that $\widehat{m}(x_i;\lambda)$ uses y_i as well as the other observations to predict y_i. This objective function is minimized when $\lambda = 0$. This problem can be avoided by using a leave-one-out estimator. Least-Squares Cross-Validation (LSCV) is the technique whereby we minimize equation 1.17, where the fit is replaced by a leave-one-out estimator

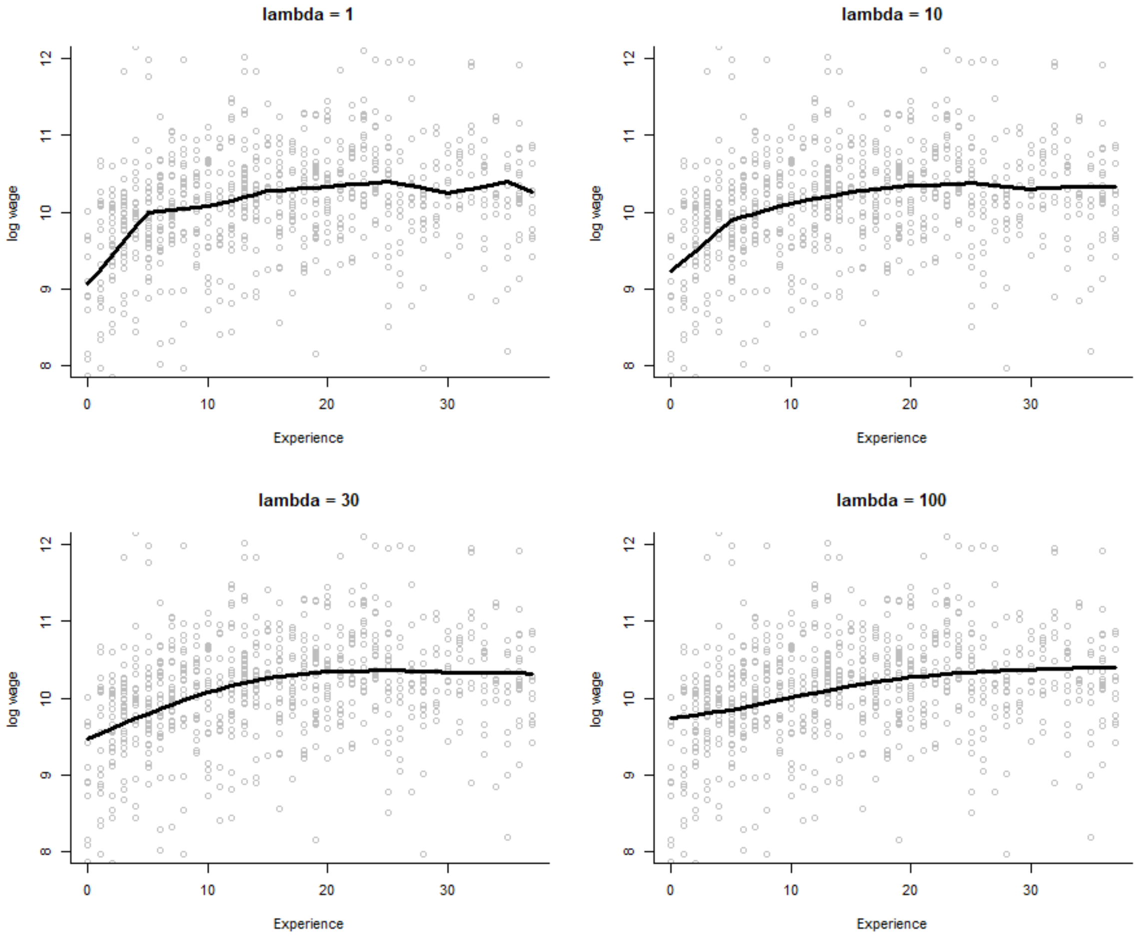
$$CV(\lambda) = \sum_{i=1}^{n} \left[y_i - \widehat{m}_{-i}(x_i;\lambda)\right]^2, \tag{1.18}$$

where $\widehat{m}_{-i}(\cdot)$ is our leave-one-out estimator, and is defined as our original nonparamet-

---

have the equivalent transformation done on the smoothing parameter $\lambda \rightarrow \alpha\lambda$ to get the same fit.

Figure 1.6: Log-wage versus Experience for College-educated Males Working in Personal Care and Service with Different Penalty (λ) Factors



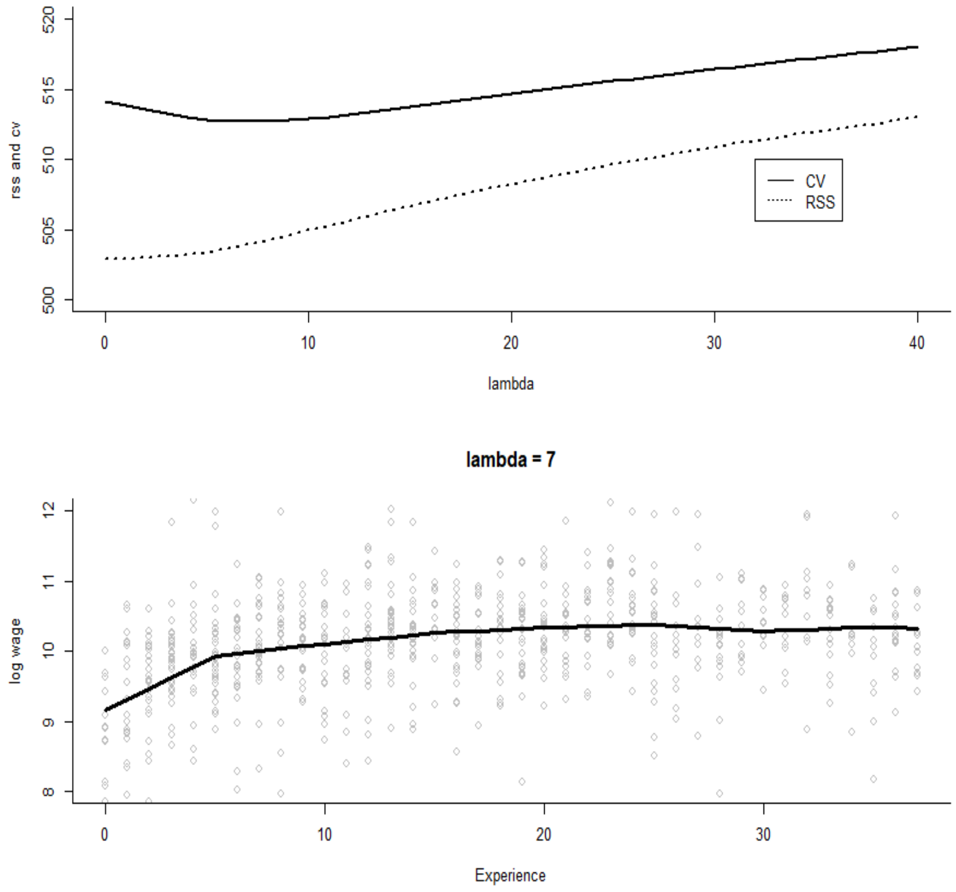
ric regression estimator $\widehat{m}(\cdot)$ applied to the data, but with the point (x_i, y_i) omitted. We will thus choose a smoothing parameter $\widehat{\lambda}_{CV}$ that will minimize $CV(\lambda)$ over $\lambda \geq 0$.

Using the same number of knots, the top panel of Figure 1.7 shows the corresponding CV and RSS curves at different values of λ. We can see that the RSS curve is strictly increasing as theory predicts and would choose a lambda of zero. The CV curve, on the other hand, decreases at first and reaches a minimum when $\lambda = 7$. The resulting fit (bottom panel of Figure 1.7) is smoother than what the RSS criterion would provide.[6]

---

[6] Note that to compute our CV statistics, we transformed equation 1.18 to avoid the high computational cost of calculating n versions of $\widehat{m}_{-i}(x_i; \lambda)$ (i.e., the order-n^2 algorithm) using fast order-n (Hutchinson and De Hoog, 1985).

Figure 1.7: Objective Functions for Choosing Penalty Factors for Linear P-splines for College-educated Males Working in Personal Care and Service



Knots and Degree Selection

Using an "optimal" lambda and CV criterion, we can compare p-spline models that use different numbers (and location) of knots and different bases (degrees). From experimenting with the number of knots and degrees, the literature finds that (1) adding more knots only improves the fit for a small number of knots (2) when using many knots, the minimum CV for linear and quadratic fits become indistinguishable. In general, we suggest using quadratic or cubic basis functions.

Though there exist more formal criterion to select the number and location of knots, Ruppert et al. (2003) provide simple solutions which often work well. Their default choice of $\mathbb{K}$ is

$$\mathbb{K} = \min\{(1/4) \times \text{ number of unique } x_i, 35\},$$

where $\mathbb{K}$ is the number of knots. For knot locations they suggest

$$\kappa_{\mathbb{K}} = \left(\frac{\mathbb{k}+1}{\mathbb{K}+2}\right)\text{th sample quantile of the unique } x_i$$

for $\mathbb{k} = 1, ..., \mathbb{K}$.

Eilers and Marx (1996, 2010) argue that equally spaced knots are always preferred. Eilers and Marx (2010) present an example where equally spaced knots outperform quantile spaced knots. The best type of knot spacings is still under debate and both methods are still commonly used.[7]

While knots' location and degree selection usually have little effect on the fit when using a "sufficiently" large amount of knots, they may become important when dealing with more complex problems. For example, when trying to smooth regression functions with strong varying local variability or with sparse data. In these cases, using a more complex algorithm to make your selection may be more appropriate.

1.3.2 Kernel and Bandwidth Selection

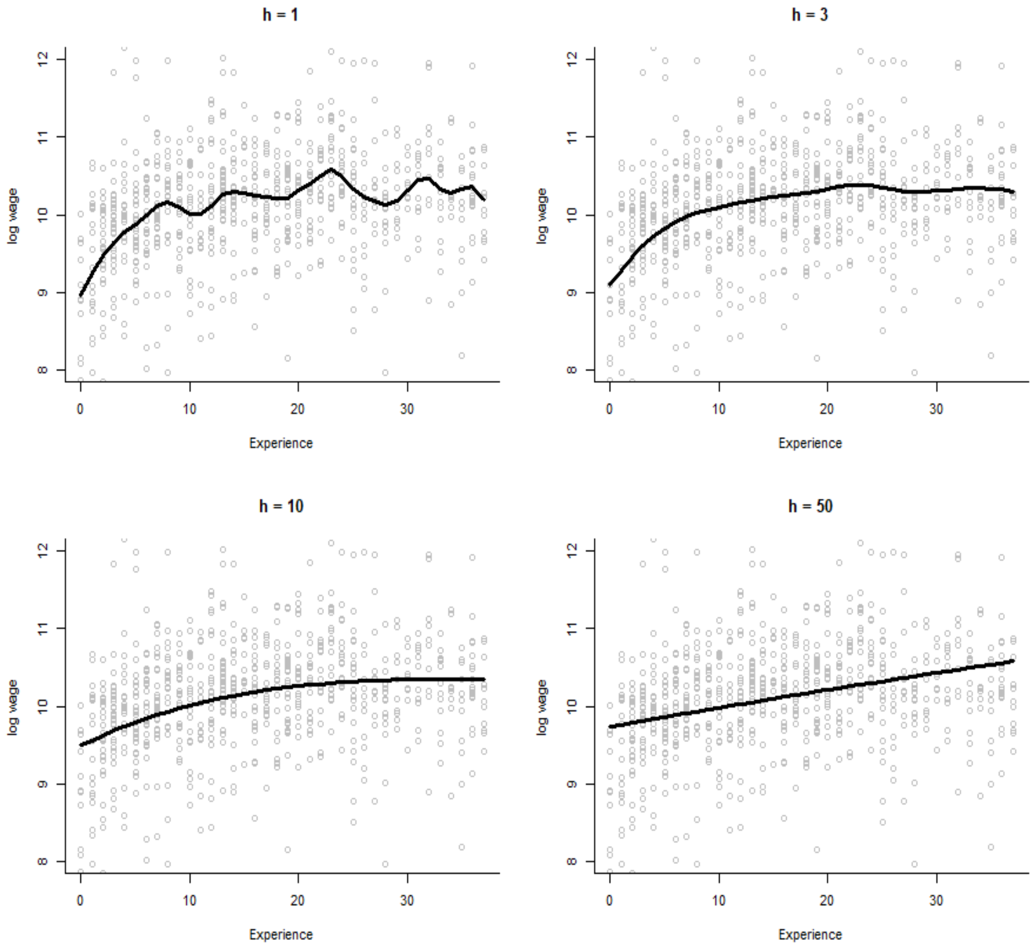
Choosing a kernel function is similar to choosing the degree of the piecewise polynomials in spline models, and choosing the size of the bandwidth (h) is similar to choosing the number and location of knots. There exist equivalents to having a direct penalty (λ) incorporated in a kernel model, but those are rarely used in applied kernel estimation. We will therefore focus our discussion on kernel and bandwidth selection.

Similar to adding more knots or decreasing the penalty λ in a spline model, decreasing the bandwidth will lead to less bias, but more variance. Figure 1.8 illustrates this effect using LLLS and a Gaussian kernel for college-educated males working in

---

[7] Montoya et al. (2014) use a simulation to test the performance of different knot selection methods with equidistant knots in a p-spline model. Specifically, they compare the methods presented in Ruppert et al. (2003) with the myopic algorithm knot selection method, and the full search algorithm knot selection method. Their results show that the default choice method performs just as well or better than the other methods when using different commonly used smoothing parameter selection methods.

Figure 1.8: Log-wage versus Experience for College-educated Males Working in Personal Care and Service when Varying the Bandwidth (h) Parameter



personal care and service. As the size of the bandwidth (h) increases, the fit becomes smoother and converges to OLS.

Choice of bandwidth and kernel can be chosen via the asymptotic mean square error (AMSE) criterion (or more specifically, via asymptotic mean integrated square error). In practice, the fit will be more sensitive to a change in bandwidth than a change in the kernel function. Reducing the bandwidth (h), leads to a decrease in the bias at the expense of increasing the variance. In practice, as the sample size (n) tends to infinity, we need to reduce the bandwidth (h) slowly enough so that the amount of "local" information (nh) also tends to infinity. In short, consistency requires that

$$\text{as } n \to \infty, \text{ we need } h \to 0 \text{ and } nh \to \infty.$$

The bandwidth is therefore not just some parameter to set, but requires careful consideration. While many may be uncomfortable with an estimator that depends so heavily on the choice of a parameter, remember that this is no worse than pre-selecting a parametric functional form to fit your data.

Cross-Validation Bandwidth Selection

In practice, there exist several methods to obtain the "optimal bandwidth" which differ in the way they calculate asymptotic mean square-error (or asymptotic mean intergrated square-error). Three typical approaches to bandwidth selection are: (1) reference rules-of-thumb, (2) plug-in methods and (3) cross-validation methods. Each has its distinct strengths and weaknesses in practice, but in this survey we will focus on the data driven method: cross-validation.[8] Henderson and Parmeter (2015) provide more details on each of these methods.

LSCV is perhaps the most popular tool for cross-validation in the literature. This criterion is the same as the one described in Section 3.1.1 to select the penalty parameter in spline regression. That is, we use a leave-one-out estimator

$$CV(h) = \sum_{i=1}^{n} \left[y_i - \widehat{m}_{-i}(x_i)\right]^2, \tag{1.19}$$

whereby we minimize the objective function with respect to h instead of λ and the (LCLS) leave-one-out estimator is defined as

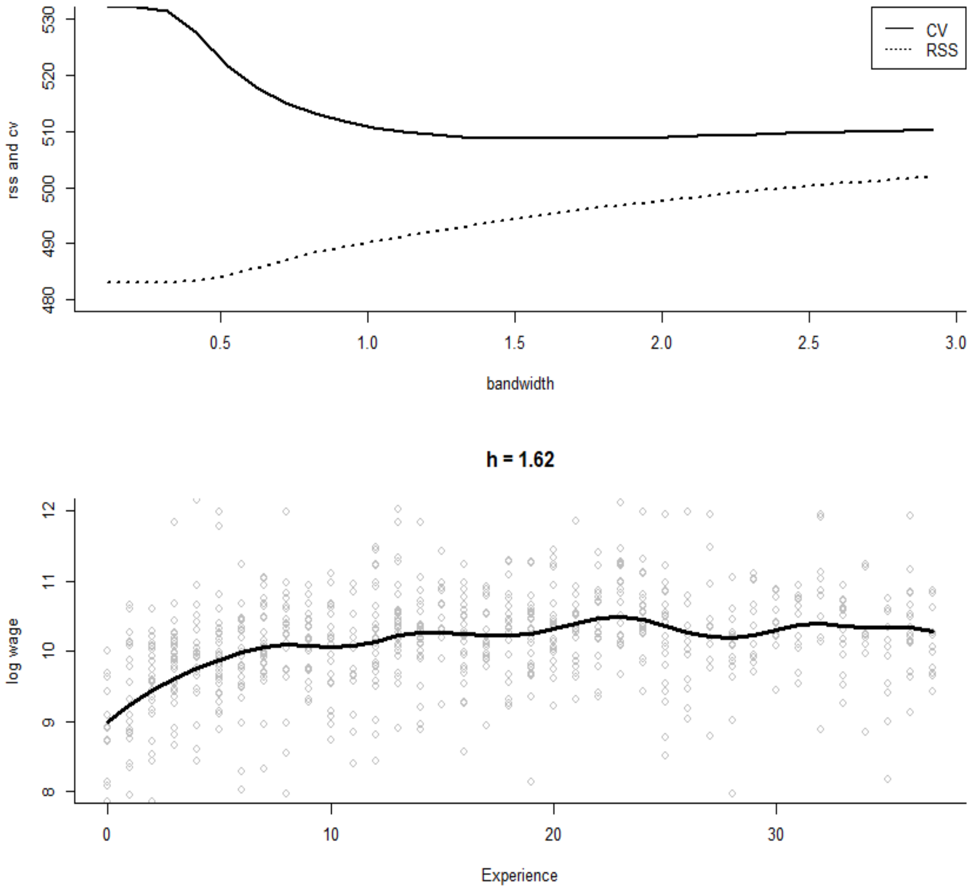
$$\widehat{m}_{-i}(x_i) = \frac{\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} y_j K_h(x_j, x_i)}{\sum\limits_{\substack{j=1 \\ j\neq i}}^{n} K_h(x_j, x_i)}.$$

In the top panel of Figure 1.9, we show an analogous figure to that presented in

---

[8] While there is no theoretical justification for doing so, it is common to use rule-of-thumb methods designed for density estimation as a form of exploratory analysis. In fact, we used a rule-of-thumb to compute the bandwidth in our previous examples (Section 2). In its general form, the bandwidth (designed for Gaussian densities with a Gaussian kernel) is $h_{rot} = 1.06\sigma_x^2 n^{-1/5}$. For the remainder of the article, we will use bandwidths selected via cross-validation.

Figure 1.9: Objective Functions for Choosing Bandwidths for Kernel Estimators for College-educated Males Working in Personal Care and Service



Section 3.1.1. It shows the corresponding CV and RSS curves for different bandwidths. When failing to use the leave-one-out estimator, the RSS curve is strictly increasing (i.e., the optimal bandwidth is zero). Using the leave-one-out estimator, the objective function is minimized at $h = 1.62$. The resulting fit, (bottom panel of Figure 1.9) shows more variation than the linear p-spline (Figure 1.7). This is not surprising as the linear p-spline forces a linear fit between each knot. The two graphs would have looked more similar if we had used a cubic p-spline, allowing for curvature between knots.

Kernel Function Selection

Kernel selection is typically considered to be of secondary performance as it is believed that it makes minor differences in practice. The optimal kernel function, in the AMISE sense, is the Epanechnikov kernel function. However, as stated previously, it may not be useful in some situations as it does not posses more than one derivative. Gaussian kernels are often used in economics as they possess derivatives of any order, but there are losses in efficiency. In the univariate density case, the loss in efficiency is around 5%. However, Table 3.2 of Henderson and Parmeter (2015) shows that this loss in efficiency increases with the dimension of the data (at least in the density estimation case). In practice, it may make sense to see if the results of a study are sensitive to the choice of kernel.

1.3.3 Splines versus Kernels

In these single-dimension cases, our spline and kernel estimates are more or less identical. Spline regressions have the advantage that they are much faster to compute. While it is uncommon to have an economic problem with a single covariate, if that were the case, we likely would suggest splines.

In a multiple variable setting, the difference between the two methods are more pronounced. The computation time for kernels increases exponentially with the number of dimensions. The additional computational time required for splines is minor. On the other hand, kernels handle interactions and discrete regressors (see Ma et al. (2015a) for using discrete kernels with splines) well (both common features in economic data). It is also relatively easier to extract gradients with kernel methods.

In reality there are camps: those who use kernels and those who use splines. However, the better estimator probably depends upon the problem at hand. Both should be considered in practice.

1.4 Instrumental Variables

Nonparametric methods are not immune to the problem of endogeneity. A first thought about how to handle this issue would be to use some type of nonparametric two-stage least-squares procedure. However, this is not feasible as there exists an ill-posed inverse problem (to be discussed below). It turns out that this problem can be avoided by using a control function approach much like that in the parametric literature (e.g., see Cameron and Trivedi (2010)).

To motivate this problem, consider a common omitted-variable problem in labor economics: ability in the basic compensation model. A correctly specified wage equation could be described as:

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 z_1 + \beta_3 abil + \epsilon \tag{1.20}$$

where $educ$ is years of education, $abil$ is ability, and z_1 is a vector of other relevant characteristics (e.g., experience, gender, race, marital status). However, in applied work, ability ($abil$) cannot be directly measured/observed.

If we ignore ability ($abil$), it will become part of the error term

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 z_1 + u, \tag{1.21}$$

where $u = \epsilon + \beta_3 abil$ and $abil$ is correlated with both u and with $educ$. Our resulting estimated return to education (β_1) will be biased and inconsistent. We can resolve this problem if we can find an instrumental variable (IV) which is uncorrelated with u (and so uncorrelated with ability), but correlated with $educ$. Several IVs have been considered in the literature for this particular model,[9] each with their own strengths and weaknesses, but for the purpose of this illustration, we will use spouse's wage. That is, we assume that spouse's wage is correlated with education, but not with ability.

In the parametric setting, the control function (CF) approach to IVs is a two-step procedure. In the first step, we regress the endogenous variable on the exogenous vector z:

$$educ = \gamma_0 + \gamma_1 z + v,$$

where $z = (z_1, spwage)$ and $spwage$ is the spouse's wage, and obtain the reduced form

---

[9] Those IVs include, but are not limited to: minimum school-leaving age, quarter of birth, school costs, proximity to schools, loan policies, school reforms, spouse's and parents' education/income.

residuals $\widehat{v}$. In the second step, we add $\widehat{v}$ to equation 1.21 and regress

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 z_1 + \beta_3 \widehat{v} + u.$$

By directly controlling for v, *educ* is no longer viewed as endogenous.

1.4.1 The Ill-Posed Inverse Problem and Control Function Approach

Let us first go back and consider the general nonparametric regression setting

$$y = m(x) + u, \tag{1.22}$$

where $E[u|x] \neq 0$, but there exists a variable z such that $E[u|z] = 0$. For the moment, assume that x and z are scalars.

Using the condition

$$E[u|z] = [y - m(x)|z] = 0,$$

yields the conditional expectation

$$E[y|z] = [m(x)|z] = \int m(x) f(x|z) dx. \tag{1.23}$$

Although we can estimate both the conditional mean of y given z ($E[y|z]$) as well as the conditional density of x given z ($f(x|z)$), we cannot recover $m(x)$ by inverting the relationship. That is, even though the integral in equation 1.23 is continuous in $m(x)$, inverting it to isolate and estimate $m(x)$ does not represent a continuous mapping (discontinuous). This is the so-called ill-posed inverse problem and it is a major issue when using instrumental variables in nonparametric econometrics.

Luckily, we can avoid this problem by placing further restrictions on the model (analogous to additional moment restrictions in a parametric model). Here we consider a control function approach. Similar to the parametric case above, we consider the triangular framework

$$x = g(z) + v, \tag{1.24}$$

$$y = m(x) + u \tag{1.25}$$

with the conditions $E[v|z] = 0$ and $E[u|z, v] = E[u|v]$. The first condition implies that z is a valid instrument for x and the second allows us to estimate $m(x)$ and avoid the ill-posed inverse problem. It does so by restricting u to depend on x only through v. More formally,

$$\begin{aligned} E[y|x, v] &= m(x) + E[u|x - v, v] \\ &= m(x) + E[u|z, v] \\ &= m(x) + E[u|v] \\ &= m(x) + r(v), \end{aligned}$$

and hence

$$m(x) = E[y|x, v] - r(v), \tag{1.26}$$

where both $E[y|x, v]$ and $r(v)$ can be estimated nonparametrically. In short, we control for v through the nonparametric estimation of the function $r(v)$.

1.4.2 Spline Regression with Instruments

Now that we have the basic framework, we can discuss nonparametric estimation in practice. Consider our previous compensation model, but without functional form assumptions:

$$log(wage) = m(educ, z_1) + u \tag{1.27}$$

where ability (*abil*) is unobserved. It is known that *abil* will be correlated with both the error u and the regressor *educ* (i.e., $E[u|\text{educ}] \neq 0$). Similar to the parametric setting, if we have an instrument z such that

$$E[u|z] = [log(wage) - m(educ, z_1)|z] = 0,$$

we can avoid the bias due to endogeneity. We again define $z = (z_1, spwage)$, where our excluded instrument, *spwage*, is the spouse's wage. This yields

$$E[log(wage)|z] = [m(educ, z_1)|z].$$

Our problem can now be written via the triangular system attributable to Newey et al. (1999):

$$educ = g(z) + v \tag{1.28}$$

$$log(wage) = m(educ, z_1) + u, \tag{1.29}$$

where $E[v|z] = 0$ and $E[u|z, v] = E[u|v]$. Similar to the parametric case, we first estimate the residuals from the reduced-form equation (i.e., $\widehat{v}$). We then include the reduced-form residuals nonparametrically as an additional explanatory variable:

$$log(wage) = w(educ, z_1, \widehat{v}) + u, \tag{1.30}$$

where $w\left(educ, z_1, \widehat{v}\right) \equiv m\left(educ, z_1\right) + r\left(\widehat{v}\right)$ and $\widehat{m}\left(educ, z_1\right)$ can be recovered by only extracting those terms that depend upon *educ* and z_1. Note that we need to use splines that do not allow for interactions between *educ* or z_1 and u (interactions between *educ* and z_1 are allowed).

In what follows, we will use cubic B-splines (the default for most `R` packages) with $\mathbb{K} = \min\{(1/4) \times$ number of unique $x_i, 35\}$ equi-quantile knots (see Section 3) in both stages. Figure 1.10 shows the fitted results for the first stage. In the left panel, we see that as individuals' experience increases, the level of education slowly decreases, with a significant drop after 35 years of work experience. In the right panel, we observe a quadratic relationship between education level and spouse's wage. That is, the higher the spouse's wage, the higher the individual's education, but for individuals whose spouse have a high level of income, the relationship becomes negative.

The fitted plots from the second stage are given in Figure 1.11. Controlling for

Figure 1.10: First-stage Estimates for Education for College-educated Males Working in Personal Care and Service versus Experience and Spousal Income

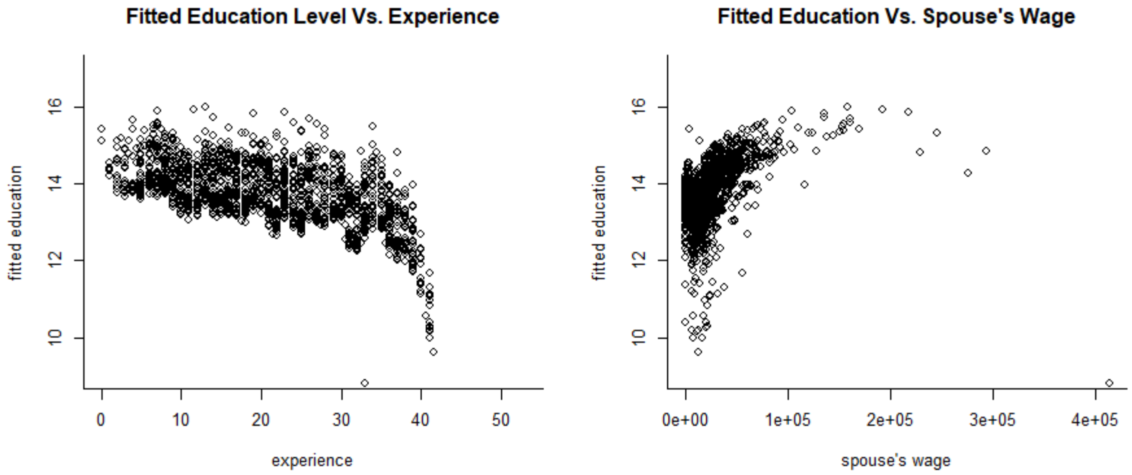
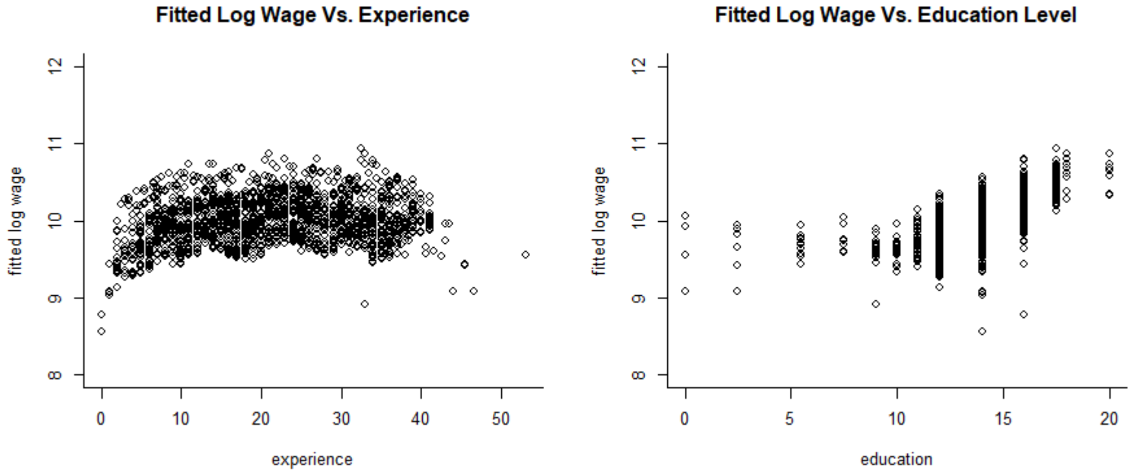


Figure 1.11: Second-stage Estimates for Log-wage for College-educated Males Working in Personal Care and Service versus Experience and Education Level



education,[10] men's log wage seems to increase in the first few years on the job, stabilizes mid-career, and then decreases towards the end of their career in personal care and service. Education seems to affect log wage positively only after 10-12 years of schooling (high-school level).

1.4.3 Kernel Regression with Instruments

Estimation via kernels is relatively straightforward given what we have learned above. We again use a control function approach, but with local-polynomial estimators.

---

[10] Recall that in our previous examples, the level of education is fixed at 16 years – college degree.

Kernel estimation of this model was introduced by Su and Ullah (2008) and is outlined in detail in Henderson and Parmeter (2015). In short, the first stage requires running a local-p_1th order polynomial regression of the endogenous regressor on z, obtaining the residuals and then running a local-p_2th order polynomial regression of y on the endogenous regressor, the included exogenous regressors and the residuals from the first stage.

More formally, our first stage regression model for our example is

$$educ = g\left(z\right) + v, \tag{1.31}$$

and the residuals from this stage are used in the second stage regression

$$log(wage) = w(educ, z_1, \widehat{v}) + u, \tag{1.32}$$

where $w\left(educ, z_1, \widehat{v}\right) \equiv m\left(educ, z_1\right) + r\left(\widehat{v}\right)$.

In spline regression, we simply took the estimated components not related to $\widehat{v}$ from $\widehat{w}\left(\cdot\right)$ in order to obtain the conditional expectation $\widehat{m}\left(educ, z_1\right)$. However, disentangling the residuals is a bit more difficult in kernel regression. While it is feasible to estimate additively separable models, we follow Su and Ullah (2008) and remove them via counterfactual estimates in conjunction with the zero mean assumption on the errors. Under the assumption that $E(u) = 0$, we recover the conditional mean estimate via

$$\widehat{m}\left(educ, z_1\right) = \frac{1}{n}\sum_{i=1}^{n} \widehat{w}\left(educ, z_1, \widehat{v}_i\right), \tag{1.33}$$

where $\widehat{w}\left(educ, z_1, \widehat{v}_i\right)$ is the counterfactual estimator of the unknown function using bandwidths from the local-p_2 order polynomial regression in the second step (derivatives can be obtained similarly, but summing over the counterfactual derivatives of $\widehat{w}\left(\cdot\right)$).[11]

Bandwidth selection and order of the polynomials (p_1 and p_2) are a little more complicated. Here we will give a brief discussion, but suggest the serious user consult

---

[11] Multiple endogenous regressors can be handled by running separate first stage regressions and putting the residuals from each of those regressions into the second stage regression and finally summing over i to obtain the conditional mean estimates.

Chapter 10 (which includes a discussion of weak instruments) of Henderson and Parmeter (2015).

Bandwidth selection is important in both stages. In the first-stage, v is not observed and we want to make sure that the estimation of it does not impact the second-stage. If we observe the conditions in Su and Ullah (2008), it allows for the following cross-validation criterion

$$CV\left(h_2\right) = \min_{h_2} \frac{1}{n} \sum_{i=1}^{n} \left[y_i - \widehat{m}_{-i}\left(educ_i, z_{1i}\right)\right]^2, \tag{1.34}$$

and the first-stage bandwidths can be constructed as

$$\widehat{h}_1 = \widehat{h}_2 n^{-\gamma}, \tag{1.35}$$

where the acceptable values for γ depend upon the order of the polynomials in each stage.[12]

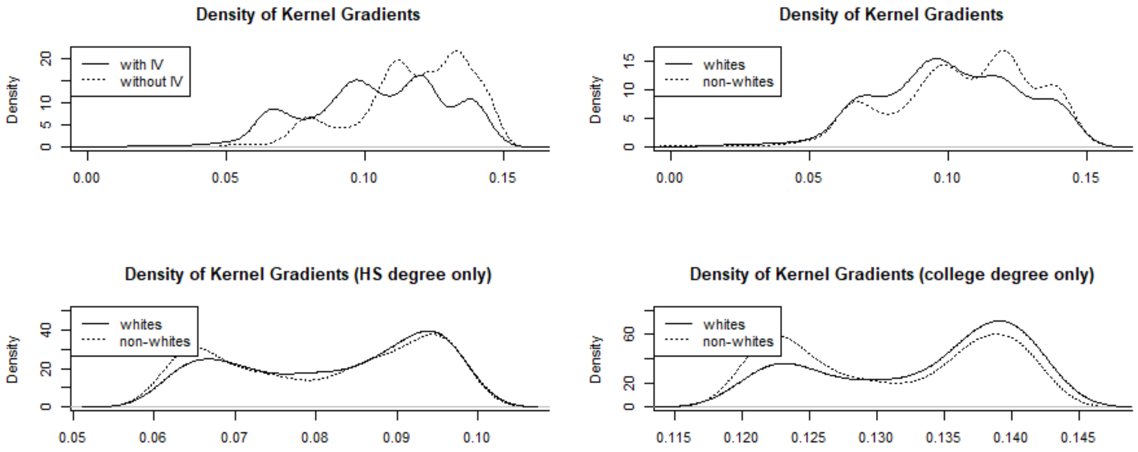
Henderson and Parmeter (2015) give the admissible combinations of polynomial orders for the Su and Ullah (2008) estimator with a single endogenous variable and a single excluded instrument. In practice, they suggest using a local-cubic estimator in the first stage (local-linear in the first stage is never viable) and a local-linear estimator in the second stage for a just-identified model with a single endogenous regressor. For other cases, the conditions of Assumption A5 in Su and Ullah (2008) need to be checked.

Using the methods outlined above, Figure 1.12 shows the impact of controlling for endogeneity. The upper-left panel gives density plots for the gradient estimates across the sample for returns to education both with and without using instrumental variables. Most college-educated men working in personal care and service have a wage increase of about 5 to 15% for each additional year they spend in school. However, the distribution is skewed to the left suggesting that a few men have seen their investment

---

[12] The acceptable range for γ is between $\left(2\left(p_2+1\right)+q_1+1\right)^{-1} \max\left[\frac{p_2+1}{p_1+1}, \frac{p_2+3}{2(p_1+1)}\right]$ and $\left(2\left(p_2+1\right)+q_1+1\right)^{-1} \frac{p_2+q_1}{q_1+q_2}$, where q_1 and q_2 represent the number of elements in the first and second stage regressions, respectively.

Figure 1.12: Second-step Gradients of Log-wage with Respect to Education for College-educated Males Working in Personal Care and Service



in education yield no returns or even negative returns (for a similar result in a nonparametric setting see Henderson et al. (2011a)).

Comparing those results with the gradients without instruments, we clearly see that failing to control for endogeneity would overestimate the returns to education (as expected). That is, the distribution of gradients without using IV is more concentrated around 10 to 15% returns and fewer low returns.

To try to swing back to the examples from before, the upper-right panel of Figure 1.12 gives densities of gradient estimates controlling for endogeneity for whites versus non-whites. The figure seems to suggest that non-whites have higher rates of return to education. This is commonly found in the literature, but is often attributed to lower average years of education. To try to compare apples to apples, in the bottom two panels we plot the densities of returns to education for fixed levels of education (high school and college, respectively). Here we see while the general shape is similar, whites tend to get more mass on the higher returns and less mass on lower returns, especially for college graduates.[13]

---

[13] We could combine spline and kernel methods to obtain an IV estimator as in Ozabaci et al. (2014). Using this combination allows for a lower computational burden and oracally efficient estimates.

CHAPTER 2
RISK-RETURN TO EDUCATION TRADE-OFFS: TO THE MEAN AND BEYOND
(WITH DANIEL J. HENDERSON AND LE WANG)

2.1 Introduction

A growing body of literature argues that, risk should be included in traditional human capital models (Heckman et al. (2008); Heckman et al. (2006); Mazza et al. (2013); Cunha and Heckman (2007)). The main concern being that, if individuals are risk-averse, a part of their estimated returns to schooling might reflect compensation for that risk. That is, individuals make decisions about their level of education the same way they make decisions about any other investment, i.e. they evaluate the risk and compare it to the expected return to determine if the investment is worthwhile. If individuals consider education to be an investment good, as opposed to a consumption good, then risk in education investments may bias estimated rate of returns upward and may be an omitted variable in the traditional human capital model.

In this paper, we examine the relationship between expected rates of return to education and higher-order risks. We want to uncover the following: Is risk compensated by a higher mean earning (i.e. is education an investment good)? Is the distribution's variance enough to capture the risk taken by individuals when they are making education investment decisions? Is risk in itself an omitted variable in the traditional human capital model? The answers to those questions will greatly depend on how we define and measure risk. Hence, we first need to clarify the definition of risk when talking about (education) investment decisions and, specifically, highlight the link between higher order risk attitudes and (education) rate of returns' distributions.

Extensive theoretical and empirical literature already exist linking individual's decision making under uncertainty with their risk attitudes. Most of the focus so far has been on 2nd order risk attitude, otherwise known as risk aversion (starting with Bernoulli, 1738). By including risk preferences (such as 2nd order risk attitude) into basic decision model, researchers have been able to better predict individuals' choices in the presence of uncertainty/risk. The main prediction of the basic 2nd order risk model is that a risk-averse individual would prefer a payoff with certainty over a risky-payoff (i.e. several options with different probability of occurring) with the same mean.

In the investment literature, those risk-preferences based predictions are the basis of portfolio theory. That is, portfolio theory assumes that investors are risk-averse, meaning they prefer a less risky portfolio to a riskier one for a given level of return. This implies that an investor will take on more risk only if he/she is expecting higher returns. Empirically, we should thus observe a positive correlation between the risk-level of a portfolio, usually measured by the variance (2nd order risk) of the distribution of its returns, and its expected return (i.e. the mean of the distribution of its returns).

However, higher order risk attitudes (HORA) may also play a key role to describe individuals' risk preferences. More recent theoretical literature on decision making under uncertainty infers that prudence (third-order risk-aversion) and temperance (fourth-order risk-aversion) also affect behavior under uncertainty. Prudence and temperance relates to individuals' aversion to negative skewness and high kurtosis, respectively (Kimball, 1990; Kimball and Weil, 1992; Eeckhoudt and Schlesinger, 2008). In particular, prudence is an aversion to what is often referred to as a "downside risk", i.e. a type of risk that decreases the skewness of a prospect but does not change its mean and variance. Similarly, temperance describes an aversion to an "outer risk", i.e. a type of risk which increases kurtosis without changing the mean, variance, and skewness of a prospect.

This would suggest that, in a risky investment decision context, individuals would require a premium (via a higher mean return) not just when facing 2nd order risk (i.e. high variance) but also when facing higher order risks such as 3rd and 4th order risk (i.e. negative skewness and high kurtosis). Thus far, HORA empirical results using various risky investments have been mixed.

While there are many different ways to think of and model education as an investment decision (with risk), our paper is most closely related to the earnings dispersion literature. This earnings dispersion literature started a while ago with Kuznets and Friedman (1939) who found a positive relationship between the mean and variance of incomes among workers from a selected few professions. Since then, many papers have found a positive relationship between the mean and variance of earnings across occupations (including Hartog and Vijverberg, 2007; McGoldrick and Robst, 1996; Cubas and Silos, 2017), across field or type of education (Christiansen et al., 2007; Tuor and Backes-Gellner, 2010), and across different levels of general education(Palacios-Huerta, 2003). Only a few of these papers also look at the relationship between skewness and the mean, and none so far looks at the kurtosis (4th order risk).

We contribute to this literature in three different ways. First, we use the distribution of returns to education (gradients) directly as opposed to the distribution of earnings or residuals. Most of the empirical literature on compensation for risk in an education context uses either the moments from the distributions of estimated earnings or the moments from the distributions of estimated residual earnings to measure risk-return trade-offs. Specifically, they estimate those earnings (or residuals) using a Mincer earnings equation, get the higher moments (variance and sometimes skewness) from the distributions of those estimates, and then add those higher moments back into the Mincer equation (Risk Augmented Mincer earnings equation). Instead of looking at the earnings dispersion, we want to look at the returns to one more year of education directly. Specifically, we use the moments from the distributions of the gradients on earnings obtained from our augmented Mincer equation. We believe that this will allow us to directly capture the risk of investing in education rather than the more general risk of the labor market captured by earning dispersion.

Second, on top of the 2nd moment (variance), we also use the 3rd and 4th moments (skewness and kurtosis) of the rate of returns' distribution to capture risk. As described previously, the theoretical literature on decision making under uncertainty suggest that higher order risk attitudes also matters when individuals make decisions. By adding those higher moments to our model, we want to investigate if there also exists compensation for 3rd and 4th moment risk-types. If we find compensation for those other types of risk then it would suggest that variance does not capture the entire risk of investing in education and that therefore those higher order risk attitudes also play a key role in education investment decisions.

Third, we both compute individuals' returns to education and estimate the relationship between mean returns and their higher moments nonparametrically. That is, we believe that both the relationship between years of schooling and earnings and the relationship between risk of investing in one more year of schooling and expected earning is non-linear (nor log-linear).

A log-linear model (as is often used) suggests that an additional year of schooling gives the same percentage increase in earnings independent of the level of education already attained and of unobservable characteristics. In reality, this is very unlikely. For example, one year of High School might not yield the same returns as one year of College. Even if we were to compare two observably identical individuals (i.e. same gender, race, educational level, etc.), they still might not get the same returns from one more year spent in school. This is due to the effect of unobservable characteristics on returns such as ability level, differences in access to credit, or differences in quality of school. Researchers have relied on different methods to account for those non-linearities and heterogeneity in returns to education: adding quadratic, cubic, quartic experience and schooling terms (Murphy and Welch, 1990), sub-sample analysis (Harmon et al., 2003b), bayesian hierarchical models (Koop and Tobias, 2004), quantile regression (Fasih et al., 2012; Martins and Pereira, 2004), and non-parametric estimation (Henderson et al., 2011b). We choose to use nonparametric estimation as it does not assume any functional form and allows for heterogeneity both across and within groups.

The relationship between risk and expected earning is also unknown and likely to be non-linear. However, the empirical literature on education investment decisions has for the most part estimated this relationship via linear or log-linear models. For example, compensation for risk may vary with the level of expected returns. Also, risk attitudes are likely to be heterogeneous which cannot be captured by one constant parameter. For example, risk attitudes might have changed over time, i.e. individuals in the 80s might be more risk-averse than individuals in the 2010s and would thus require more compensation for risk (have different gradients). Our nonparametric model allows us to observe this heterogeneity which so far has been ignored in the literature.

We estimate both the rates of return to education and compensations for risk nonparametrically using categorical regression spline. In particular, for every quinquennial year between 1980 and 2015, we regress nonparametrically the log wages on education and other control variables to obtain returns to education (gradients) for each individual. We then split the sample by quinquennial year and state to obtain, for each sub-sample, the first four respective moments. Finally, we regress nonparametrically the mean on variance, skewness, and kurtosis, where each higher-moment is capturing a different type of risk.

Following the theoretical literature's predictions, we expect to find positive gradients on variance and kurtosis, and negative gradients on skewness. Specifically, we expect risk-averse individuals to seek compensation for a wider spread, i.e. bigger variance. We also expect them to have a preference for positively skewed distribution (i.e. to be prudent), and hence seek compensation for negative skewness. The former stems from the idea that prudent individuals will prefer a distribution where the unlikely events are on the right-end tail - a really high rate of return (positive skewness has a long right-end tail) - rather than a distribution where the unlikely events are on the left-end tail - a really low rate of return (negative skewness has a long left-end tail). Preferences on kurtosis are a little less obvious, but generally speaking, we expect individuals with temperance preferences to seek distributions with lighter tails (high kurtosis - leptokurtic type of distribution) where they are more likely to end up close to the mean and median rather than distributions with heavier tails (low kurtosis -

platykurtic type of distribution) where they are more likely to end up in the tails (far away from the mean).

Overall, we find positive and statistically significant compensations for variance. 60% and 80% of the compensations for skewness and kurtosis respectively are statistically significant. Risk, as measured by the higher moments of the rates of return distributions, thus plays an important role in the educational decisions of individuals. Hence, education seems to be treated as an investment good. Additionally, both skewness and kurtosis are found to have a non-linear relationship with expected returns, displaying the importance of using a nonparametric model. Interestingly, we find that most of the gradients on skewness are positive. Considering that most of our distributions have positive skewness (due to the nature of returns to education), this result means that individuals seek compensation for positive skewness in the context of educational investment.

The paper is organized as follows. In the next section we present our methodology. We then describe our data. In section 4, we present and analyze our empirical results. Finally, we conclude and discuss policy implications.

2.2 Methodology

To estimate if there are risk-return to education trade-offs, we first estimate returns to education for each individual. Hence, our model consists of four main steps: (1) estimate individuals' rate of returns to education, (2) add those rates to our original dataset, (3) split the sample by year and state and obtain our various measures of risk, and (4) establish if there is indeed compensation for that risk.

In our first step, we use the Mincer equation which regresses log wage on education and experience as a basis. Specifically, we regress nonparametrically the log of individuals' weekly wage on education and other control variables for each of the 8 years in our sample separately:

$$ln(wage_{it}) = m(educ_{it}, X_{it}) + \epsilon_{it} \tag{2.1}$$

for individuals i and years $t = 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015$; and where $wage$ is the weekly wage, $educ$ corresponds to years of education starting from 1st grade, and X is a matrix of all our control variables (i.e. years of experience, age groups, and dummies for marital status, children living at home, race, topcoded for high earnings, metropolitan area, and regions).

For computational time reason, we favor spline regression over kernel to regress our model nonparametrically. However, spline regression cannot deal with dummy variables as is. We therefore opt to use categorical regression splines (CRS), a method developed by Ma et al. (2015b), which combines both spline regression for continuous variables and kernel regression for categorical variables. That is, in our case, both education and experience will be regressed using splines while the rest will be regressed using kernel. According to Ma et al. (2015b)'s simulations results, for a small amount of continuous variables (less than 5) and a large sample size (more than 500), as we have here, CRS would have a lower computational burden compared to a kernel-only regression. Moreover, their simulations show that CRS outperforms the only two spline regression alternatives in the presence of categorical variables: a frequency estimator which breaks the sample into subsets and an additive regression spline model (if additivity is not fully present).

As we are facing potential endogeneity problems, we use Horowitz (2011)'s approach which allows the use of instrumental variables in a nonparametric context. Specifically, we use the CRS R package developed by Nie and Racine (2012) (recently updated: Racine et al. (2018)) which includes this IV approach using CRS. We choose this approach over the other proposed IV method included in the CRS package, Darolles et al. (2011)'s approach, for computational speed reasons: the latter is computationally more costly when using large datasets like ours.

Horowitz (2011)'s approach is analogous to the IV approach in the linear case (OLS with IV) and thus have similar requirements for its instruments to be valid. That is, we add the following restriction to equation (2.1):

$$E(\epsilon|W = W_{it}, X = X_{it}) = 0 \tag{2.2}$$

where W is an instrument for education (i.e. *educ*). It follows that, using the CRS methodology, we estimate the function $m()$ such that

$$E(ln(wage_{it}) - m(educ_{it}, X_{it})|W_{it}) = 0$$

Similarly to the parametric IV, we want an instrument W that is correlated with years of education, *educ* but not with ability (and thus not affect weekly wage directly). Unlike the regular linear IV approach, however, the number of instruments cannot exceed the number of endogenous variables. As we only consider education to be endogenous in our model, we can only use one instrument.

There already exist an extensive literature on the role of ability in the wage equation and which instrument(s) could potentially be used to control for that endogeneity. All of those instruments have been shown to have some limitation. Some of the most recurring IVs include cost of schooling in terms of money and time (Card, 2001; Card, 1993), changes in legislation/reforms (Harmon and Walker, 1995; Meghir and Palme, 2005), birth date/cohort (Angrist and Krueger, 1991; Angrist and Krueger, 1992; Kaymak, 2009), combination of birth and reforms (Acemoglu and Angrist, 1999), combination of birth and family background (Winters, 2015), and family background (Blackburn and Neumark, 1993; Parker and Van Praag, 2006; Hoogerheide et al., 2012). We use family background as our instrument as it is the only one available in our CPS dataset for all time periods used. Specifically, our dataset includes parents' education, spouse's education, parents' income, and spouse's income (bottom of Table 2.1). As our model allows for only one instrument, we use spouse's yearly wage which has the least amount of missing values.

From our t (8 years) regressions[1], we obtain gradients for each individual. Indi-

---

[1] Note that all of our model parameters, i.e. our choice of degree, segments (number of knots) and

viduals' rate of return to education will simply be their respective gradient on education which is defined as:

$$rr_{it} = \frac{\partial ln(wage_{it})}{\partial educ_{it}} \tag{2.3}$$

We then add those back into our original dataset (step 2).

In our third step, we split our full sample by year t and state j. Then, we obtain the first four (4) moments from each subsample:

$$Mean_{tj} = r\bar{r}_{tj};$$

$$Var_{tj} = \frac{1}{N_{tj}} \sum_i (rr_{itj} - r\bar{r}_{tj})^2;$$

$$Skew_{tj} = \frac{1}{N_{tj}} \sum_i (rr_{itj} - r\bar{r}_{tj})^3;$$

$$Kurt_{tj} = \frac{1}{N_{tj}} \sum_i (rr_{itj} - r\bar{r}_{tj})^4.$$

In words, "Var" is simply the variance of each subsample. "Skew", the skewness of each subsample, measures the degree and direction of asymmetry in the distribution. "Kurt", the kurtosis of each subsample, measures the heaviness of the tails. For example, a normal distribution has a kurtosis of 3, we would thus consider anything above 3 to be a distribution with relatively "light" tails and high peak and anything below 3 a distribution with relative "fat" tails and low peak. N_{tj} is the sample size for each of the 408 subsamples (=51 states times 8 years).

In our final step, we use spline regression[2] to regress the mean of returns to education on its corresponding higher moments (our measure of risk) for each subsample:

$$Mean_{tj} = g(Var_{tj}, Skew_{tj}, Kurt_{tj}) + u_{tj} \tag{2.4}$$

We also regress our final step nonparametrically, because we think that the relationship between risk and return to education might be nonlinear. First, a risk-averse individ-

---

knots' location for splines and of lambda (bandwidths) for kernel, were selected via cross-validation.

[2] A simpler and computationally faster spline regression is possible here, as we only have continuous variables.

ual may require different level of compensation at different level of risk. That is, that individual may not act the same way at the margin in a relatively low-risk environment compared to a high-risk environment. Second, individuals may have heterogeneous risk attitudes. That is, relatively more risk-averse individuals may require larger earnings compensation for the same level of risk than relatively more risk-seeking individuals. Some of that heterogeneity may be capture by subsample analysis. For example, one may infer that some occupations may be more prone than others to attract risk-seeking individuals, hence, sub-sampling by occupation may capture some of that effect. However, even if we sub-sample our analysis further we will still not be able to capture all the risk-attitudes' heterogeneity. That is, two individuals within the same occupation might still yield different risk-attitudes and thus require different compensation for the same level of risk.

Following the theoretical literature on decision making under uncertainty, we expect to see a risk-return to education trade-off, that is, we expect our gradients on variance and kurtosis to be positive and our gradients on skewness to be negative.

2.3 Data

We use individual-level data obtained from the March Current Population Survey (CPS) for quinquennial survey years between 1980 and 2015. The CPS provides information on individual characteristics such as earnings and educational attainment in the calendar year preceding the March survey. To avoid special circumstances, we focus our analysis on men in the labor force with a positive income between the age of 20 and 59 (working age) living in a household (excludes group quarters). We also exclude self-employed men as they tend to have unreliable reported income and a higher risk tolerance (Cramer et al., 2002; Ekelund et al., 2005). Our final sample is composed of over 245,000 men which can be distinguished by race, age, years of education (0 to 20), occupations, household characteristics (children and marital status dummies), and geographical location (state dummies, region dummies, and metropolitan area dummy).

Following the standard Mincer earning equation we use log weekly wages as our dependent variable[3], after adjusting for inflation. The 69 individuals with zero weekly wage have undefined log and were therefore not included in the regressions. We control for top-coded wages using a dummy.

In our first step, we control for the following variables. Education, our main variable of interest, is simply the number of years spent in school starting from 1st grade. As we do not have information about individuals' work experience, we estimate potential years of experience as *exp = age - years of schooling - 6* (negative numbers were replaced by zero). To avoid multicollinearity issues by having both age and experience together in our model, we use age groups instead. The age group (ordered) variable *ageg* is equal to 1 for men in their 20s, 2 for men in their 30s, 3 for men in their 40s, and 4 for men in their 50s. Information about race was combined into one dummy, which equals to 1 if the individual is not white. We also control for household characteristics: children and marital status. The *child* dummy is equal to 1 if one or more children are living in the same house. The *married* dummy is equal to 1 if the individual is both married and living with its spouse, and equal to 0 if the individual is married but the spouse is absent (not living at home), or if the individual is divorced, widowed, or has never been married. Finally, we also control for regional effects using nine regional dummies and for city effects using a metropolitan (*metro*) dummy which equals to 1 if the individual lives in a central city or right outside a central city.

We use our state variables to separate our samples in the third step of our analysis. That is, we infer that each state's labor market faces different risk due to different "local" conditions and laws. We then obtain their respective risk as described in the methodology section.

From the descriptive analysis of our full sample (Table 2.1), we observe that the average worker earns a wage of about \$830 per week, has about 13 years of schooling (starting with 1st grade), has 19 years of work experience, and is about 39 years old. Also, 13.8% of our sample is non-white, 67.5% is married (with the spouse present in the household), and 55.7% has at least one child living with them. Note that the last

---

[3] We use weekly wage instead of yearly wage to control for productivity

Table 2.1: Descriptive Statistics

Statistic	N	Mean	St. Dev.	Min	Max
survey year	245,401	1998	11.820	1980	2015
occupation	245,401	14.838	8.010	1	26
lnwkwage	245,332	6.474	0.732	−3.510	10.159
weekly wage	245,401	830.469	744.062	0.000	25,813.310
education	245,401	13.393	2.867	0	20
experience	245,401	19.153	10.798	0	53
married	245,401	0.675	0.468	0	1
age	245,401	38.545	10.584	20	59
age_20's	245,401	0.242	0.428	0	1
age_30's	245,401	0.301	0.459	0	1
age_40's	245,401	0.266	0.442	0	1
age_50's	245,401	0.191	0.393	0	1
non-white	245,401	0.138	0.345	0	1
topcoded	245,401	0.011	0.106	0	1
child	245,401	0.557	0.497	0	1
metro	245,401	0.603	0.489	0	1
new-england	245,401	0.088	0.283	0	1
middle atlantic	245,401	0.119	0.324	0	1
east-north central	245,401	0.136	0.343	0	1
west-north central	245,401	0.100	0.300	0	1
south atlantic	245,401	0.165	0.371	0	1
east-south central	245,401	0.048	0.213	0	1
west-south central	245,401	0.092	0.289	0	1
mountain	245,401	0.112	0.315	0	1
pacific	245,401	0.142	0.349	0	1
father's yearly wage	15,460	38,963.97	34,284.40	−22,947.71	629,097.90
spouse's yearly wage	175,498	17,175.33	23,264.22	0.00	781,440.00
father's education	15,460	11.973	3.53	0.00	20.00
mother's education	20,436	11.71	3.33	0.00	20.00
spouse's weekly wage	131,621	487.93	517.24	0.00	29,784.59

five variables in Table 2.1 are the five possible instrument variables (IVs) we considered. However, too few individuals reported their father's income, father's education, and mother's education, for them to be used in our regressions. The IV we use is therefore the spouse's wage. On average, the spouse of the men in our sample earn \$17,175 per year (about \$488 per week). We also considered using the given spouse's weekly wage but favored yearly wage as it was reported at a slightly higher rate[4].

In our analysis, we regress each year separately. While the full sample has between 24,112 and 36,655 individuals per year, we use a random sample of 8,000 individuals for every year to improve computational speed and comparability between years/regressions.

2.4 Results

Our analysis is composed of the four following (4) steps: (1) first we regress using categorical regression splines with IV (Ma et al., 2015b) for each year and get estimates of the rate of returns to education of each individual in our sample; (2) second we add and save the estimated rate of returns to education back into our original dataset, (3) third we split the sample by year and state, obtain the higher moments (mean, variance, skewness, kurtosis) for each year-state rate of returns' distribution; (4) fourth we regress the mean (1st moment) of returns to education on their higher moments to see if we observe compensation for each different type of risk.

2.4.1 Step 1: Estimated Rates of Returns

In our first step, we regress using CRS the log weekly wage on schooling and other control variables (described in the data section) for every quinquennial year. The summary statistics of men's gradients on schooling for each year are presented in Table 2.2[5]. For example, the first row of Table 2.2 represents the summary statistics of the returns to schooling for all the men in our sample surveyed in 1980. Hence, on average,

---

[4] Note that though 165,646 of the men in our sample are married, 175,498 reported their spouse's income. This difference is due to men reporting their spouse's income while their spouse is not living under the same roof.

[5] The R-squared for each of those regresions ranged from .3 to .65

men surveyed in 1980 earned an additional 7.37% on their weekly wage for one more year spent in school. During that year, one individual had a negative return to education of -471%, the lowest in our full sample.

Table 2.2: Return to Education (Step 1) Summary Statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1980	-4.7091	0.066572	0.075297	0.073695	0.086621	1.966512
1985	-0.25829	0.075467	0.092269	0.101121	0.127339	0.444304
1990	-0.94173	0.154395	0.223064	0.216229	0.29036	1.439623
1995	-0.35442	0.143902	0.179078	0.18874	0.236433	0.535378
2000	-2.632	0.20061	0.2449	0.246531	0.299516	4.13339
2005	-1.37279	0.249255	0.280718	0.271562	0.30783	1.099607
2010	-0.37826	0.134026	0.184908	0.175276	0.224808	0.68207
2015	-1.49372	0.148785	0.229941	0.221728	0.29253	1.50259

For comparison purposes, we also regressed for every year the same regressors using OLS (adding polynomial terms for education and experience) without and with an IV (spouse's wage). We found estimated rate of returns that ranged from about 6% to 12% in the model without IV, and from about 4% to 23% in the model with IV. While the estimated rate of returns using the polynomial model (with IV) are always different from the nonparametric model, the trend is overall similar. Whereas the OLS without IV estimated rates of return do not increase as much over time. Those results suggest that the often used OLS coefficients are within the rates of return to education distributions found via nonparametric estimation.

Overall, we note a general increase in individuals' returns to education over time. The two panels in Figure 2.1 show this trend over time. The top panel plots every individual's gradient on schooling for each year. The upper and lower lines respectively connect the maximum gradient for each year together and the minimum gradient for each year together. This shows us how the range of returns to education varied over time. The middle line connects the mean gradients together. The bottom panel zooms in on the mean gradient to better show its evolution over time. The two dashed lines represent the mean of the 95% coverage upper and lower bounds of the gradients on schooling. Those upper and lower estimates are a "close" alternative to the commonly used standard errors in parametric models. Apart from a general increase throughout

Figure 2.1: Gradients on Schooling (Education) from Step 1

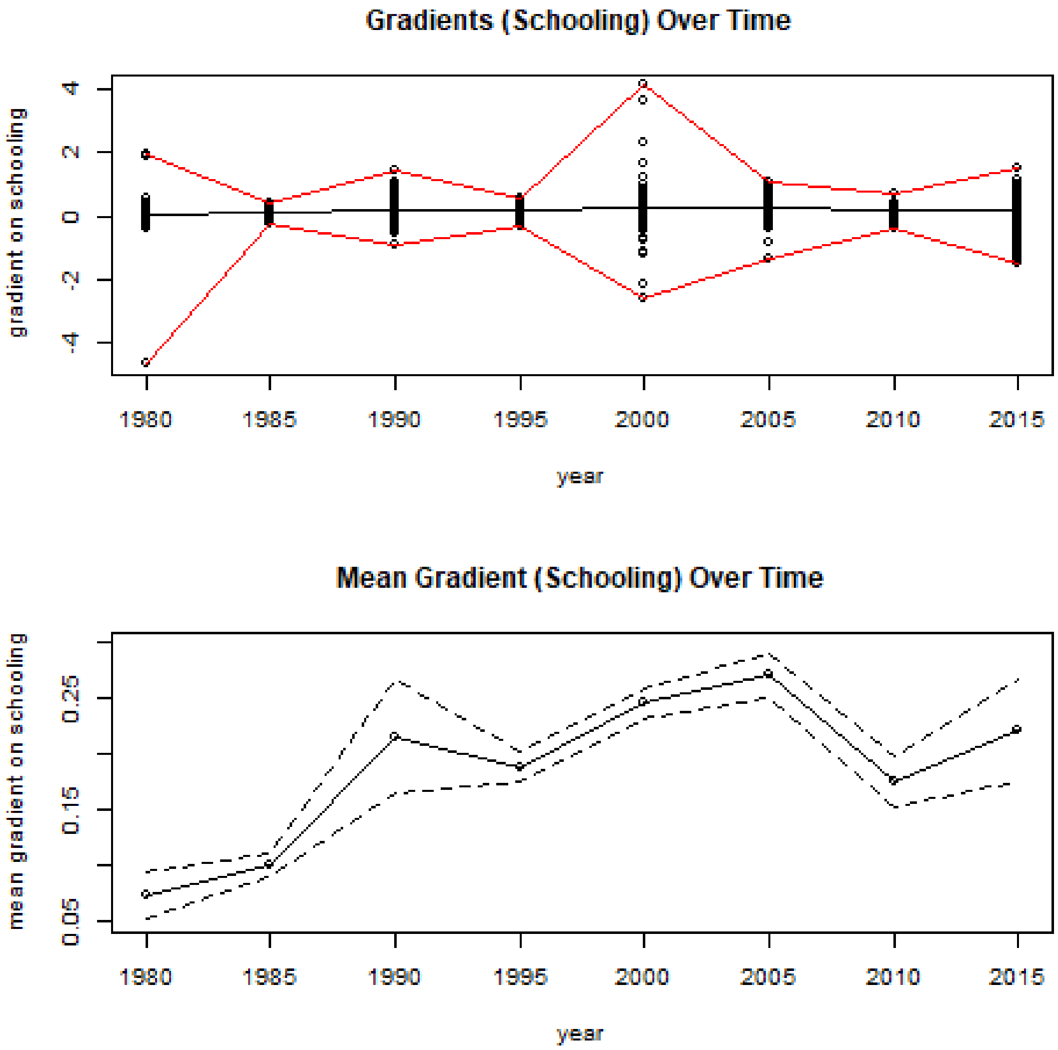
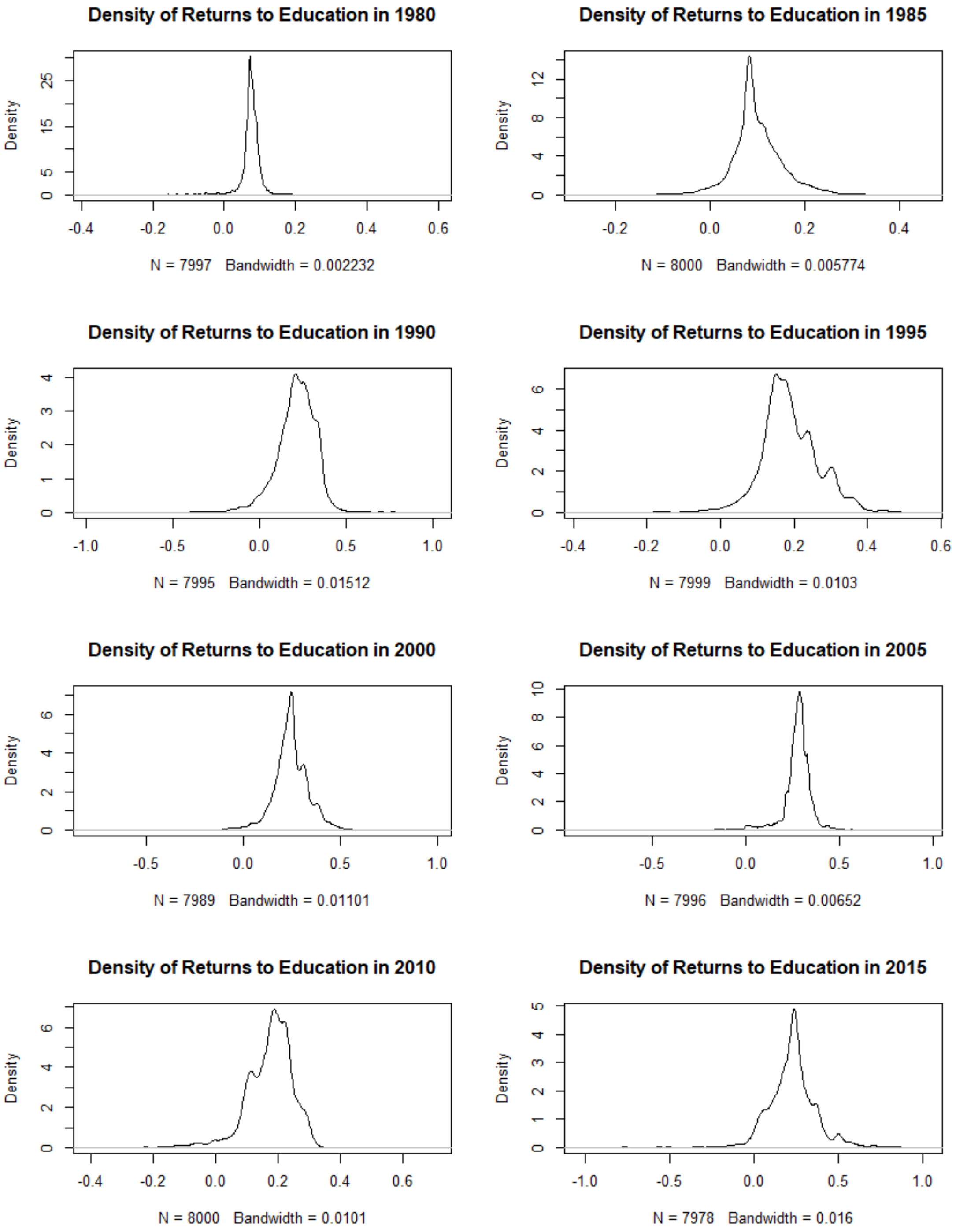


Figure 2.2: Densities of Gradients on Schooling (Education) from Step 1



the whole period, we note a significant drop in mean returns in 2010, the aftermath of the financial crisis, followed by a slight increase in 2015.

Figure 2.2 shows the resulting density of rate of returns to education for each year. Note that those distributions do not appear to be normal distributions. Apart maybe from 1980, they all seemed to be skewed in some way and to varying degree. Their variance and tails (kurtosis) also seem to vary from year to year.

2.4.2 Step 2 and 3: Moments by Year and State

In our second step, we add those gradients back into our original dataset, hence matching each individual's rate of return to education with their other characteristics. We then split our estimated rates of return to education by states as well as years (step 3). From this we get a different distribution of returns to education for each states within each year in our sample[6]. We then compute their respective four moments: mean, variance, skewness, and kurtosis. We thus get 408 observations (=51 states times 8 years) for each moment. Table 2.3 presents the summary statistics for each of the 4 moments obtained. To have a better idea of what the data looks like, we also plotted the density of each moment in Figure 2.3.

Table 2.3: Moments Summary Statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
mean	0.06	0.14	0.20	0.19	0.24	0.29
variance	0.01	0.03	0.05	0.05	0.07	0.28
skewness	-0.17	0.00	0.01	0.02	0.02	0.76
kurtosis	0.00	0.00	0.00	0.02	0.01	2.98

Table 2.4: Moments Correlation Matrix

	mean	variance	skewness	kurtosis
mean	1.00	0.86^{***}	0.25^{***}	0.06
variance	0.86^{***}	1.00	0.55^{***}	0.46^{***}
skewness	0.25^{***}	0.55^{***}	1.00	0.83^{***}
kurtosis	0.06	0.46^{***}	0.83^{***}	1.00
	Note:	*p<0.1;	**p<0.05;	***p<0.01

Table 2.4 shows the correlation between each moment. There is a positive correlation between expected returns (mean) and all three higher moments. However,

---

[6] The plots of those distributions are available upon request

Figure 2.3: Moments Density

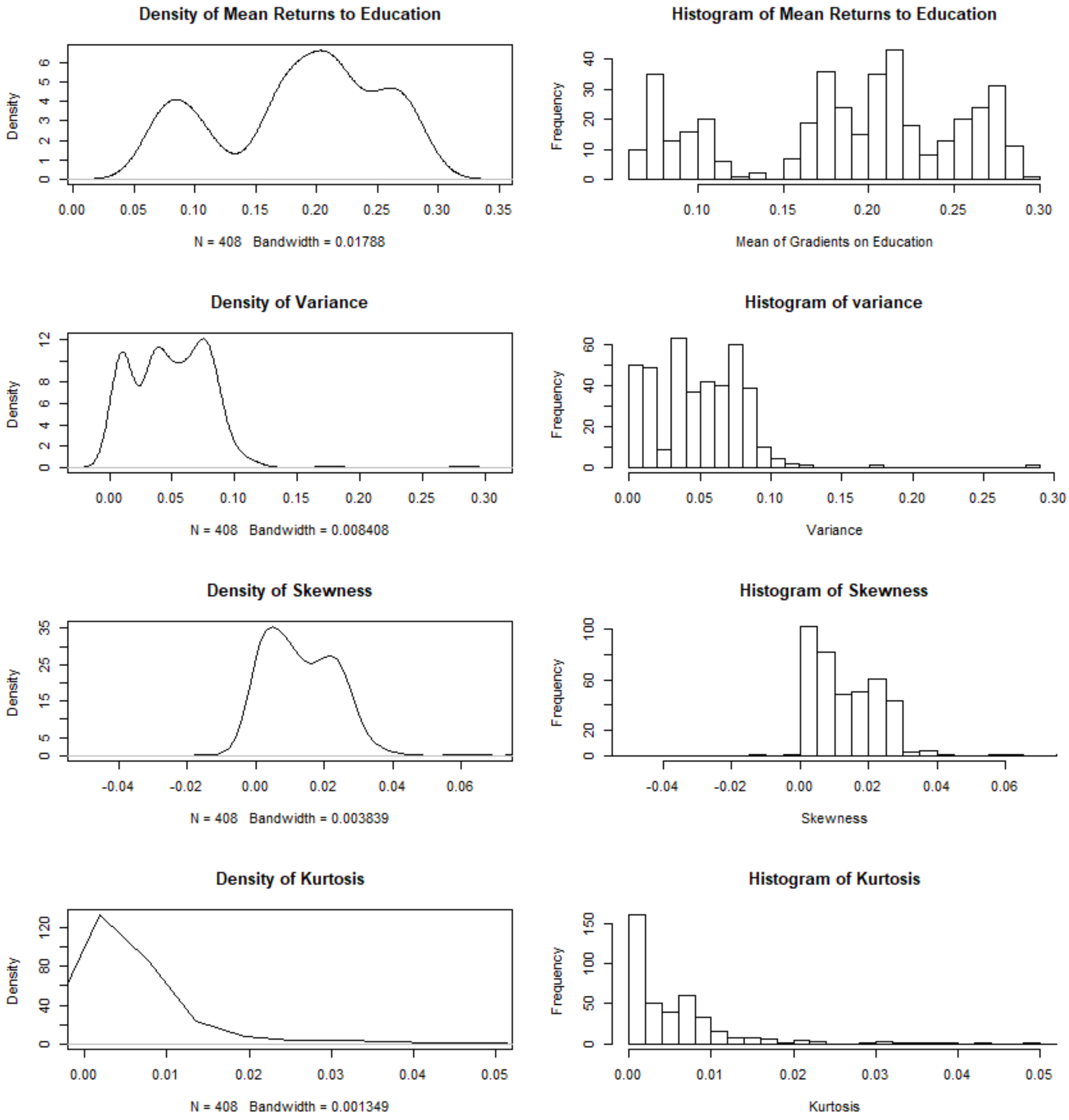
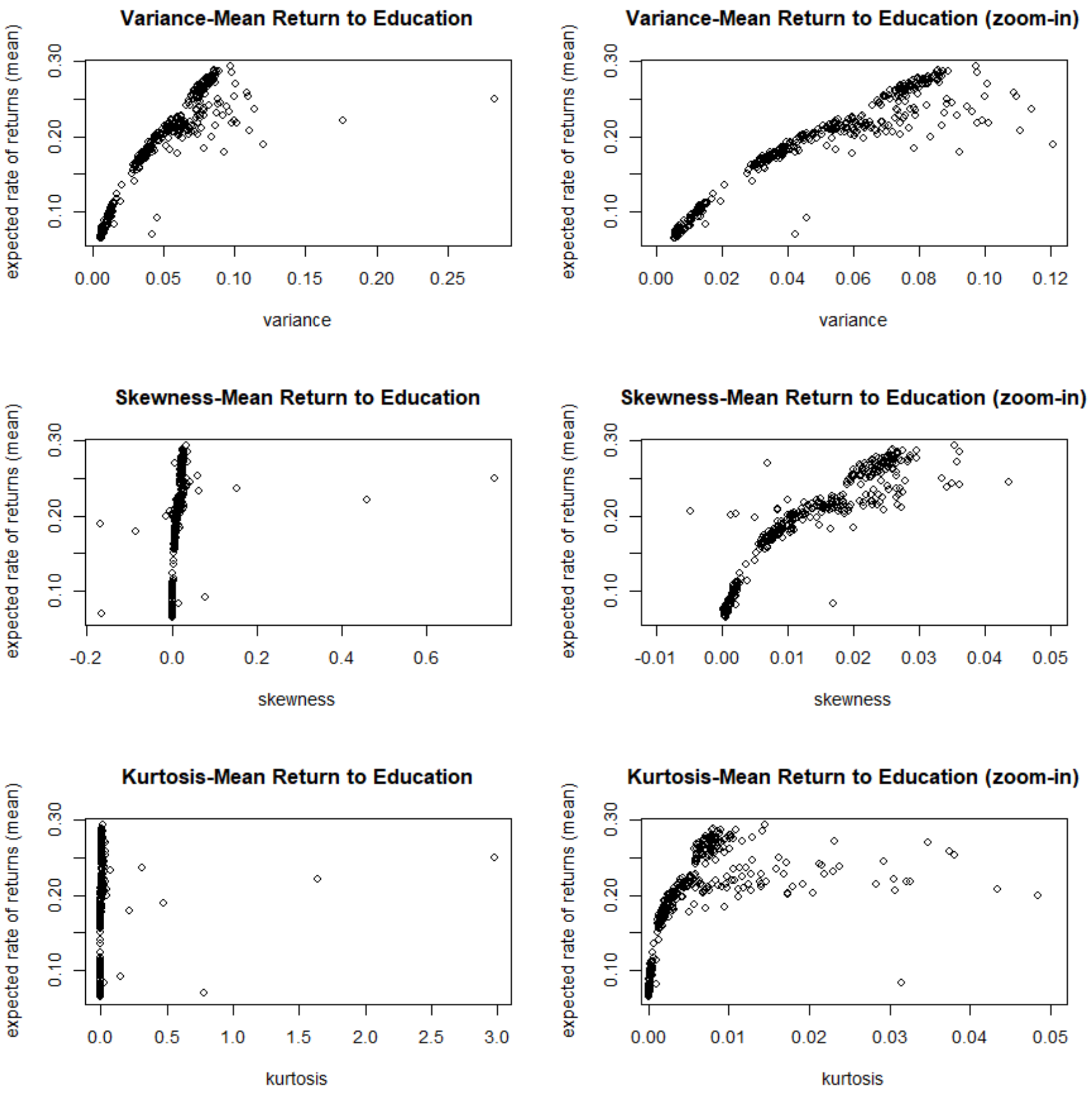


Figure 2.4: Higher Moments Against the Expected Returns (Mean)



the correlation with the 4th moment, kurtosis, is low and not statistically significant. While the correlation of expected returns with variance and with skewness are both statically significant, the correlation with skewness is much lower (0.25).

Figure 2.4 plots the higher moments against the expected rate of returns to schooling (mean). For each row of plots, the right panel is only a zoom-in of the left panel (i.e. excludes extreme values). Plotting the expected returns against their variances (top two panels) yields an unsurprising result: the higher the variance of the distribution of rate of returns to schooling, the higher the mean seems to be.

The second row of plots in Figure 2.4 shows the relationship between the skewness and the mean of each distribution. The majority of the 408 distributions are positively skewed. Considering that income has a lower bound (i.e. zero), it is not surprising that rates of return distributions tend to be positively skewed. The more positively skewed the distribution is (i.e. the right tail is longer: some extreme high rates of returns), the higher the expected returns seems ot be. This could be indicating that individuals see any type of skewness (positive or negative) as a risk and seek compensation for both types.

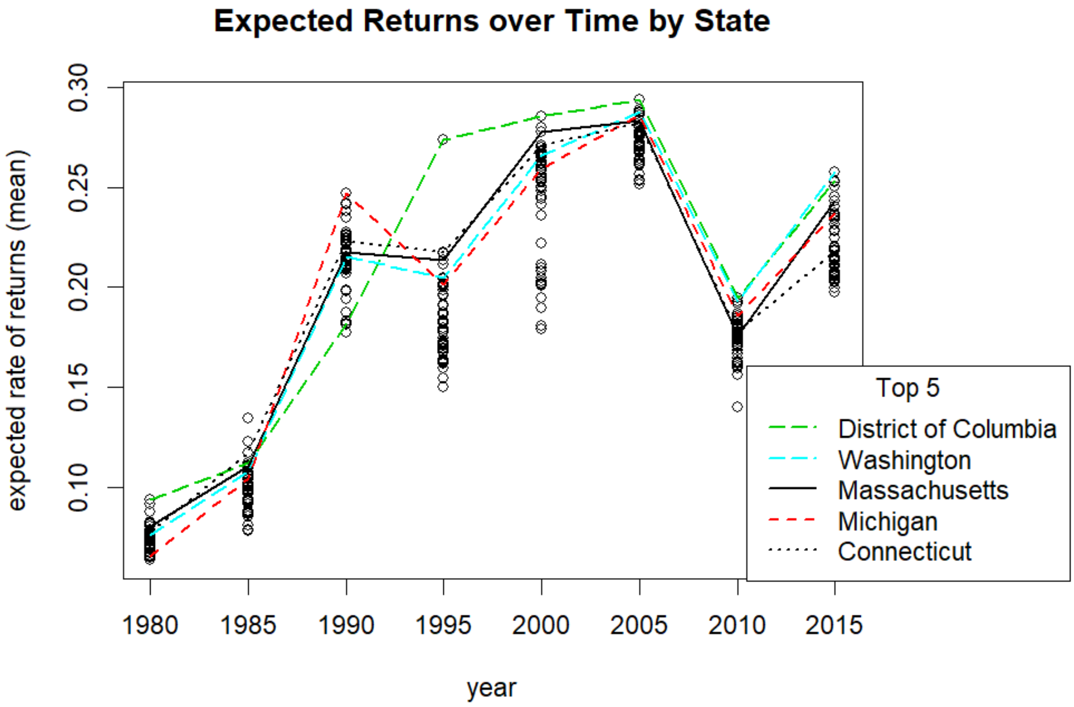
The last row of panels in Figure 2.4 plots the kurtosis against the expected rate of returns to schooling for each of the state-year labor markets. While the left panel only shows us that most of the distributions have a kurtosis close to zero, the right panel seems to indicate a positive relationship between the 1st and 4th moments of the distributions for very low level of kurtosis (below 0.01). Remember that a normal distribution has a kurtosis equal to 3. Any distribution with a kurtosis under 3 is called a platykurtic distribution, where "platy-" means "broad". Relative to a normal distribution, a platykurtic distribution has thinner tails (i.e. flat and wide distribution). Looking at the bottom-right panel of Figure 2.4, we note that the kurtosis in our sample are all between 0 and 3 meaning they are all platykurtic distributions. While a few are higher than .5, most are very close to zero, that is, have very thin tails.

Also, interestingly, risk seems to be correlated with time (year) i.e. the variance is rising over time. The correlation between the variance and year is of about 0.6[7]. This contradicts the results of Harmon et al. (2003a)'s paper on the dispersion of rate of returns to education. They obtained their variance using a random coefficient model and found that the variance in rate of returns has not increased over time. They link this result to the rapid expansion in participation in higher education and argue that if the expansion was driven by easier access to credit and by dipping further into the ability distribution then we should observe an increase in the variance of returns over time. While their results do not support this explanation, ours do.

---

[7] The correlations between the other two higher moments and time are much lower. Skewness increases slightly over time, i.e. becomes more positive, with a correlation of 0.13 with year. Kurtosis, however, vaguely increases over time with a correlation of .02 with year.

Figure 2.5: Expected Returns Top 5 and Bottom 5



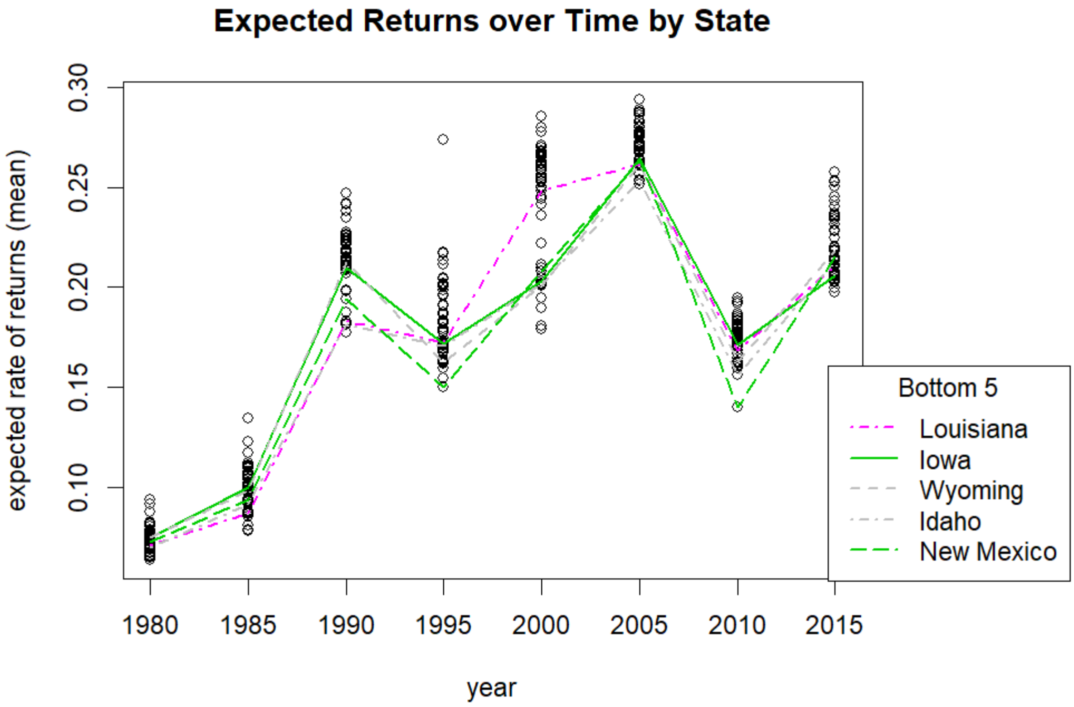
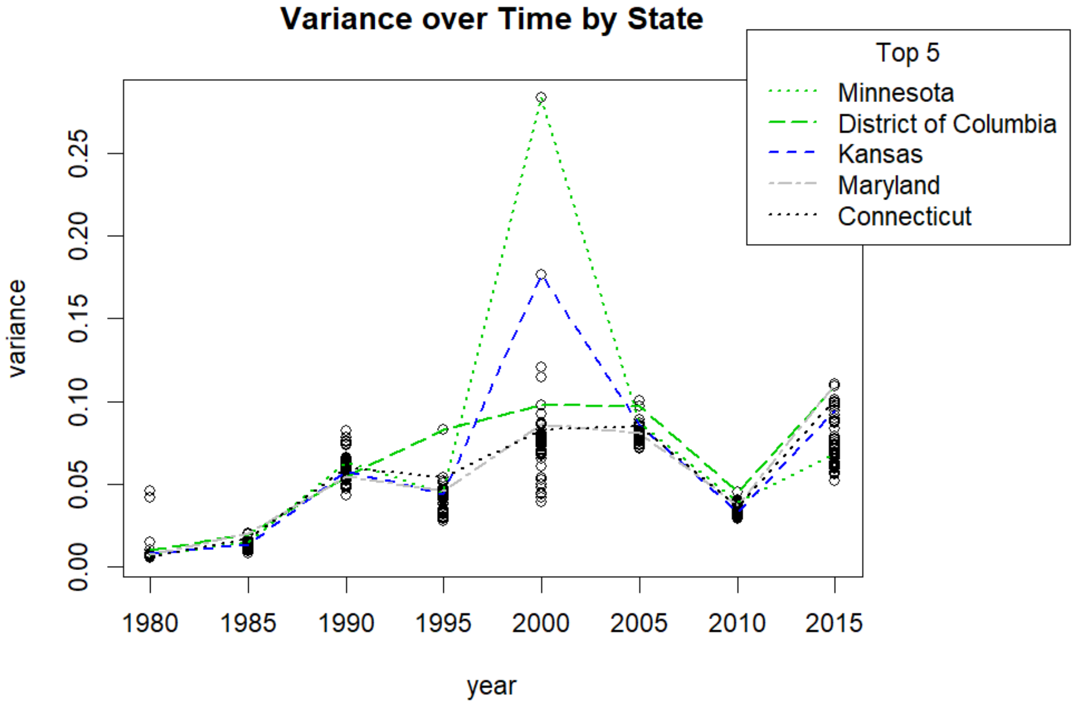


Figure 2.6: Variance Top 5 and Bottom 5



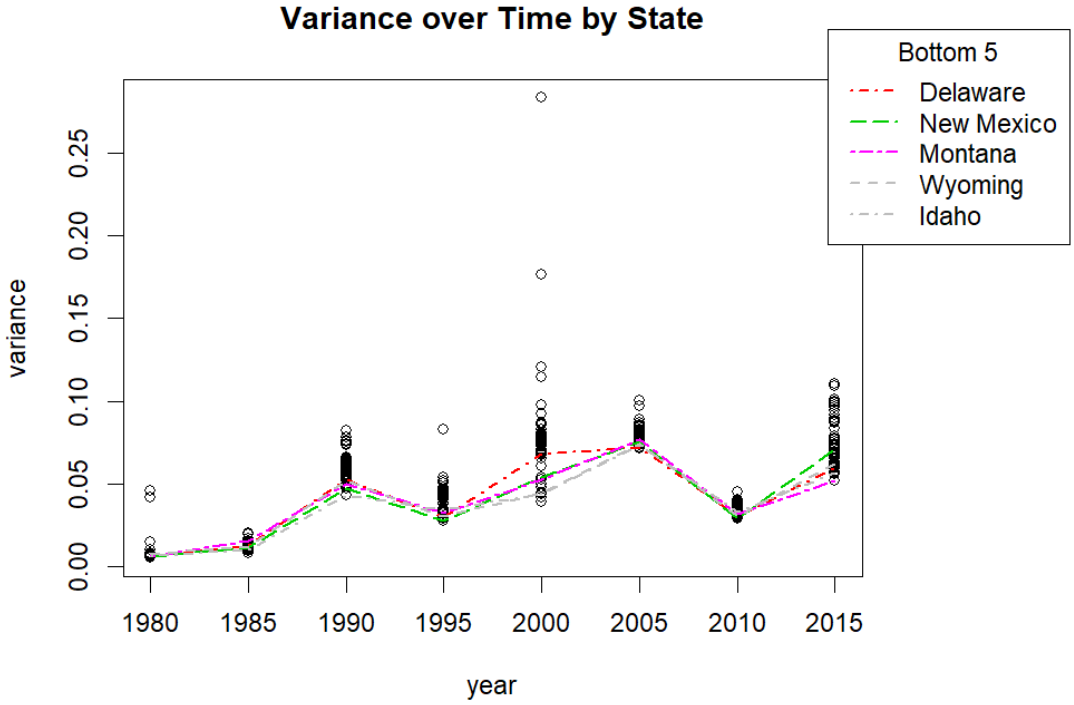


Figure 2.7: Skewness Top 5 and Bottom 5

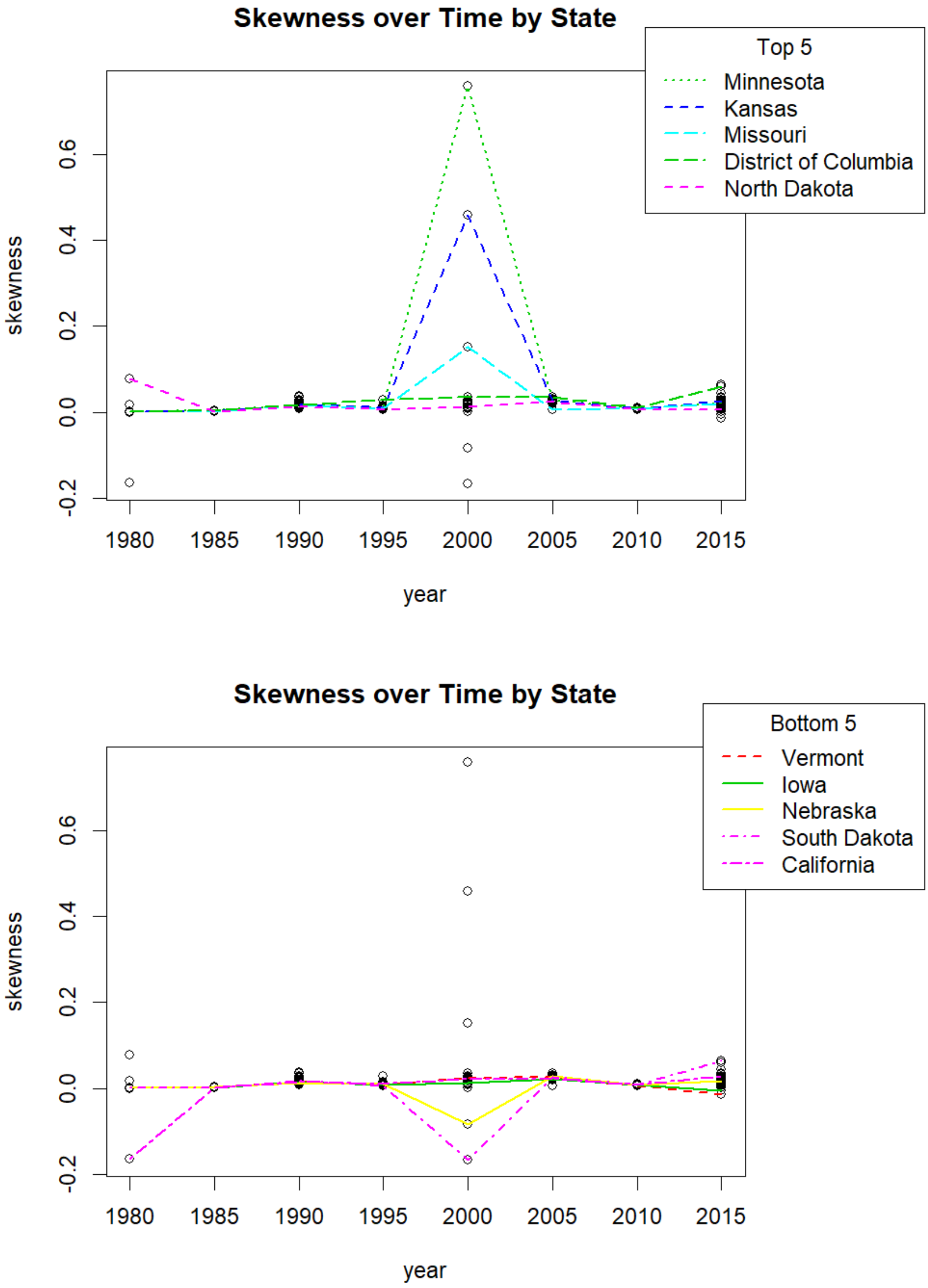
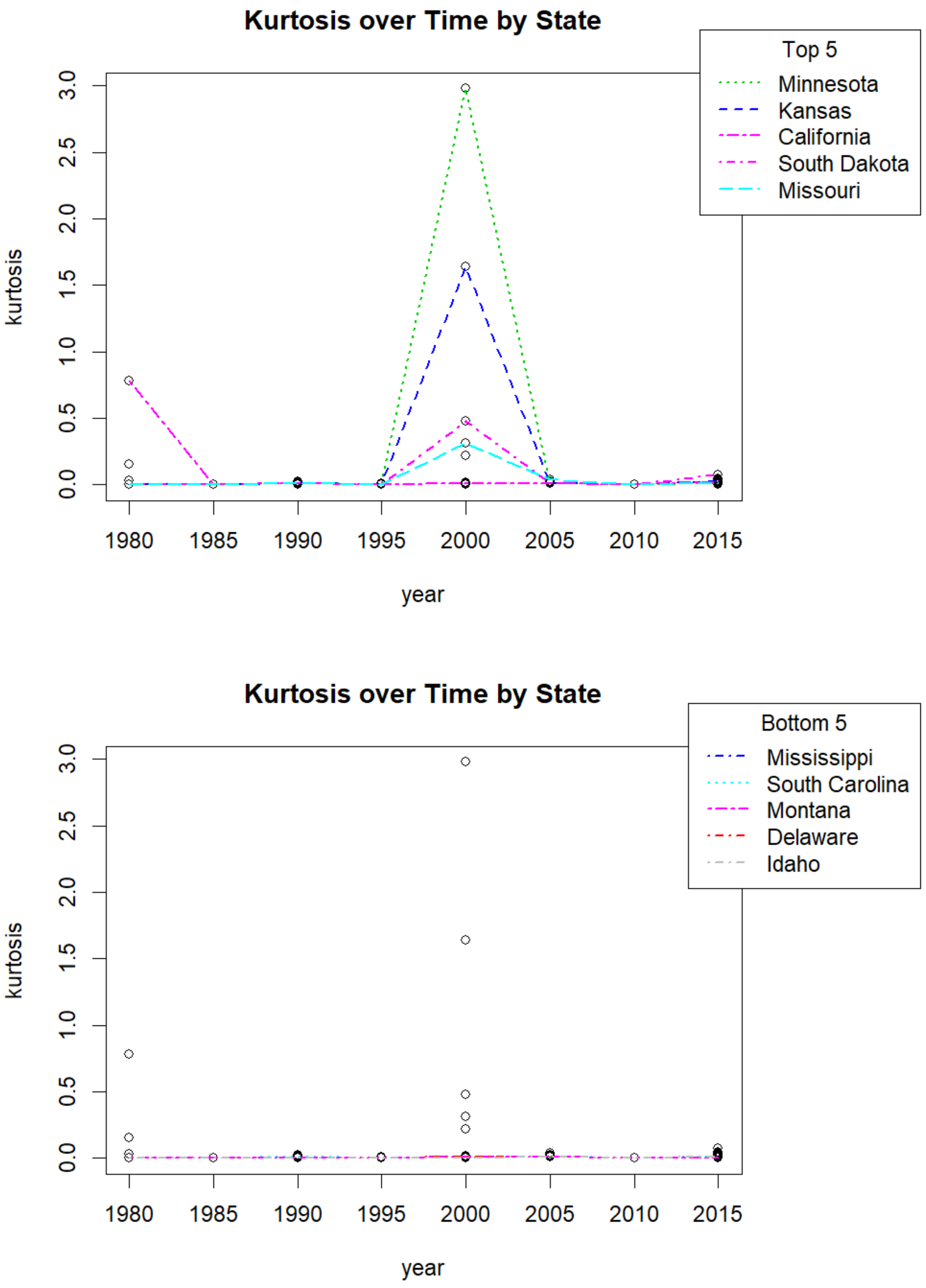


Figure 2.8: Kurtosis Top 5 and Bottom 5



Figures 2.5 - 2.8 show how each moment varies over time with each dot representing one state-year distribution. Each figure highlights the top 5 (top panel) and bottom 5 (bottom panel) states over the whole period. For example, the District of Columbia (DC) had on average the highest expected returns over the period (figure 2.5: top panel) while New Mexico had the lowest expected returns (figure 2.5: bottom panel).

While expected returns have been generally increasing over time, figure 2.5 shows that they have been decreasing both between 1990 and 1995 and between 2005 and 2010. Those two decrease could be attributed to the small recession of 1990 and the economic crisis of 2008 respectively[8]. The range of expected returns were particularly large in 2000 and 1995 indicating higher than usual inequality between states.

Figure 2.6 shows a slight increase in the variance over time. Note that overall, the variance increased when expected returns increased and decreased when expected returns decreased. Note that a high variance indicates high inequality in rate of returns to education within each state. While the year 2010 had unusually high inequality between states, it also had high variance on average compared to other years indicating high inequality within states as well. Minnessota and Kansas, in particular, stand out with very high variance in 2000. Those two states also stand out with very high skewness and kurtosis in 2000 (figures 2.7 and 2.8).

Figures 2.7 and 2.8 show skewness and kurtosis over time respectively. While both do not vary by much from year to year, the years 1980 and 2000 both display a few states with extreme values[9]. This suggests that the risk faced by individuals was particularity unequal between states. Note that Minnesota and Kansas stand out again with extremely high skewness and kurtosis in 2000.

---

[8] Note that DC was an exception to the early 1990s drop in expected returns. Its particularly low expected return to education in 1990 (compared to other states) was followed by a dramatic increase in its expected return between 1990 and 1995.

[9] While 2015 does not have particuarly extreme values, it does show a wider range than usual for both skewness and kurtosis

2.4.3 Step 4: Risk-Return Trade-Offs

In our final step (4), we regress using spline regression the mean of return to education on its higher moments (i.e. variance, negative skewness, and kurtosis) to investigate if there is indeed compensation for risk. We also add a time trend and states fixed effect[10]. Our model specification were found via cross-validation[11]. Note that the relationship between expected returns and variance is found to be linear. However, the relationships between expected returns and the higher moments skewness and kurtosis are found to be nonlinear.

Tables 2.5 and 2.6 present the summary statistics of the gradients found via spline regression. In Table 2.6, we exclude gradients that are not statistically significant. We use the 95% confidence bounds of our estimates to define statistical significance. That is, a gradient is found to be not statistically significantly different from zero if zero is included between its upper and lower bounds. While all of the gradients on variance are found to be statistically significant, about 40% and 20% of the gradients on skewness and kurtosis respectively are not statistically significant. As a lot of our distributions in our sample have kurtosis very close to zero (with litte variation), it is not surprising that the gradients on those very low kurtosis are not found to be statistically significant. The gradients on skewness found to be not statistically significant are also on small values of positive skewness and are all negative. This results is to be expected as there should not be any compensation for symmetric distributions (i.e. skewness close to zero).

For comparison purposes, we also show the OLS results using the same moments found nonparametrically in our first step. Compared to the spline results, OLS seems to overestimate the compensation on variance while underestimating the compensation on skewness (0.15 is less than the 1st quartile of the skewness gradients 0.85). More

---

[10] Note that adding states fixed effect does not seem to matter as the results are unchanged if not included

[11] We use cross-validation with the standard crs function restriction to find the model specifications for our model without controlling for time and state effects. To be able to compare this model with the one with time and state effects, we used the same model specifications (i.e. degree of 1 for all moments and segments of 1, 10, and 10 for variance, skewness, and kurtosis respectively. Note that those model specifications and results are very close to those found using cross-validation for the model including time and state effects.)

Table 2.5: Spline Gradients Summary Statistics

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
g.variance	1.34	1.34	1.34	1.34	1.34	1.34
g.skewness	-4.74	-1.38	0.04	0.33	3.04	4.28
g.kurtosis	-69.79	-5.35	-1.07	0.74	9.81	51.02

Table 2.6: Spline Gradients Summary Statistics: Statistically Significant Results Only

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	obs.
g.variance	1.34	1.34	1.34	1.34	1.34	1.34	408
g.skewness	-0.77	0.85	3.04	2.05	3.23	4.28	245
g.kurtosis	-69.79	-5.35	-1.07	0.23	9.81	51.02	326

strikingly, the OLS coefficient on kurtosis seems to suggest that a higher kurtosis will result in a lower expected return which contradicts theory. However, the spline results find gradients of much higher magnitude and of both signs suggesting a much more complex relationship between kurtosis and expected return.

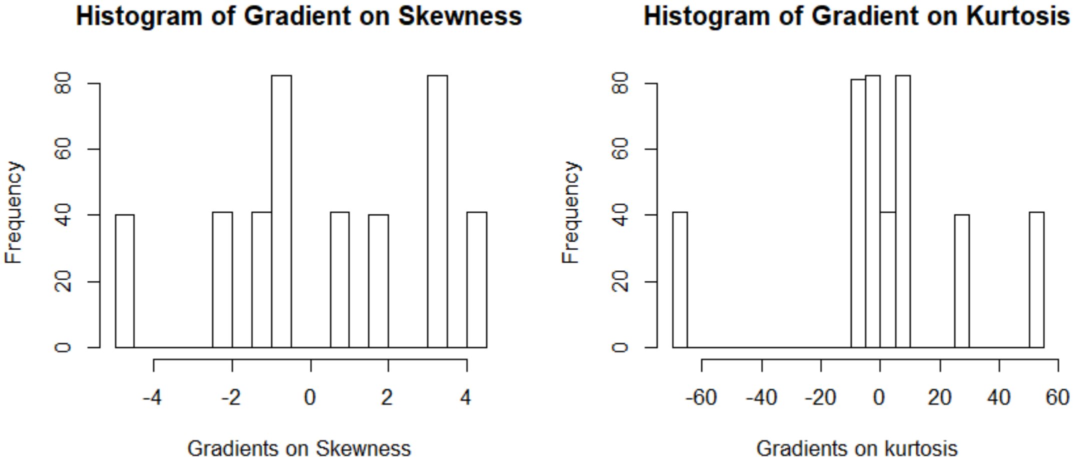
While Figure 2.9 shows the histograms of gradients on skewness and kurtosis, Figures 2.10- 2.13 plot the gradients of each higher moment against the first moment (left panel) as well as against their respective moment (right panel). Note that we do not plot this for the variance as there is no variation in the gradients of variance across both the expected returns and the variance itself (as the relationship is linear). Figure 2.14 shows the statistically significant gradients on 45° plots with their respective 95 percentile upper and lower bounds.

The gradients on variance are all positive, as expected, and statistically significant. That is, we do find compensation for 1st order risk as has been found in the literature. Interestingly, we find that this relationship is linear and thus constant over both the expected returns and the variance themselves. The gradients are equal to 1.34. This means that for any level of variance, increasing the variance by 1 unit will increase the expected return by 1.34 unit on average. However, the variance ranges from 0.01 to 0.28, therefore an increase of 0.1 units or 0.01 units would be more reasonable than an increase of 1 unit. Hence, if variance increases by 0.01 units, the expected rate of returns to education will increase by 0.0134 units or 1.34 percentage points. Going, for

Table 2.7: OLS Results (Step 4)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.08	0.01	10.36	0.00
variance	2.17	0.06	38.91	0.00
skewness	0.15	0.05	3.29	0.00
kurtosis	-0.19	0.01	-16.09	0.00
time	0.00	0.00	2.14	0.03

Figure 2.9: Spline Gradients Histograms (Step 4)



example, from 20% to 21.34% which is not a negligible amount.

The effect of skewness on the expected returns is a bit more complicated, and shows a relationship that could not have been captured by a simple linear model. First of all, due to its nature educational investment is quite unique compared to other types of investments in that rate of returns to education distributions (and wage distributions as well) are mostly positively skewed. As we are dealing with mostly positively skewed distributions, it is less clear what type of results we should expect. However, we know that the empirical literature on educational investments has found that skewness either has no effect or that the relationship is negative, implying that the more positively skewed a distribution is the lower its expected returns is going to be on average.

Our results show that all of the distributions with a negative skewness (upper left quadrant on the right panel of Figure 2.10 and 2.11) have positive gradients. This suggest that negative skewness is not compensated by a higher expected rate of return to education as we expected. The gradients on the negatively skewed distributions are

Figure 2.10: Spline Gradients on Skewness from Step 4

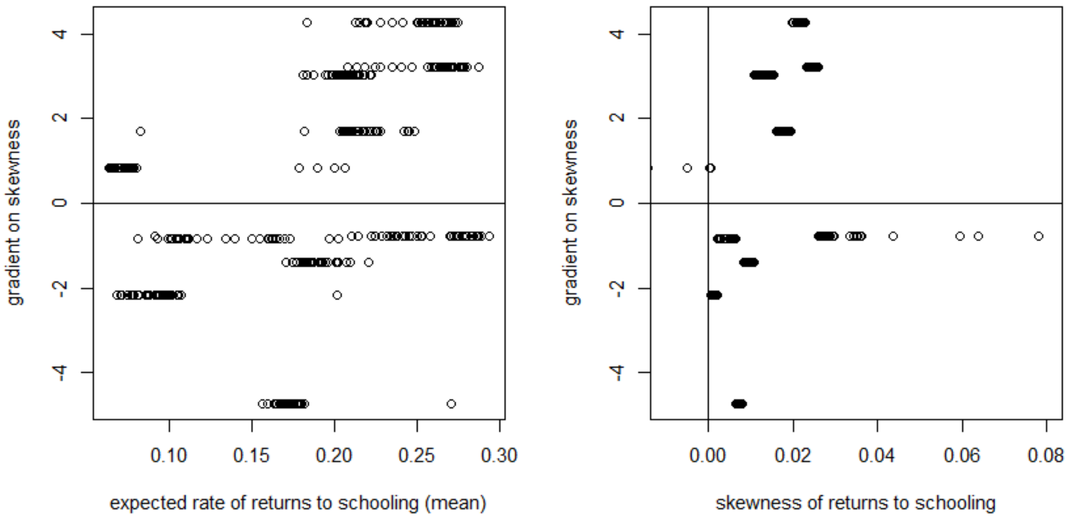


Figure 2.11: Spline Gradients on Skewness from Step 4: Stat. Sign. Only

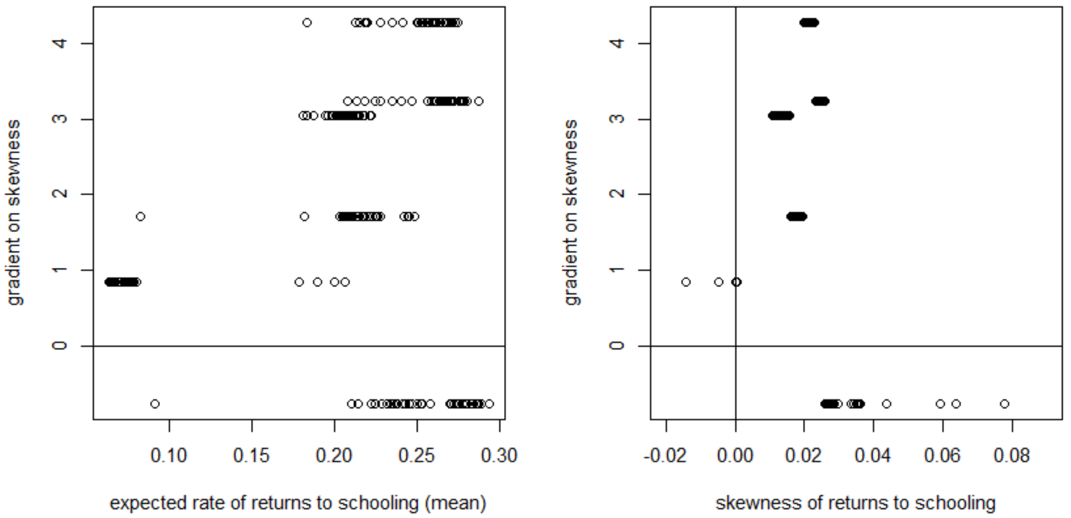


Figure 2.12: Spline Gradients on Kurtosis from Step 4

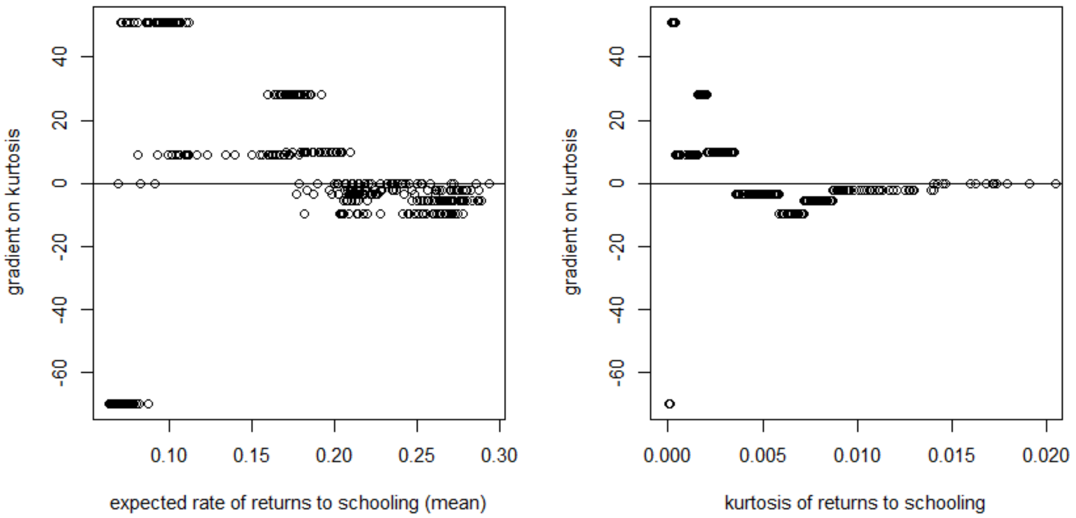


Figure 2.13: Spline Gradients on Kurtosis from Step 4: Stat. Sign. Only

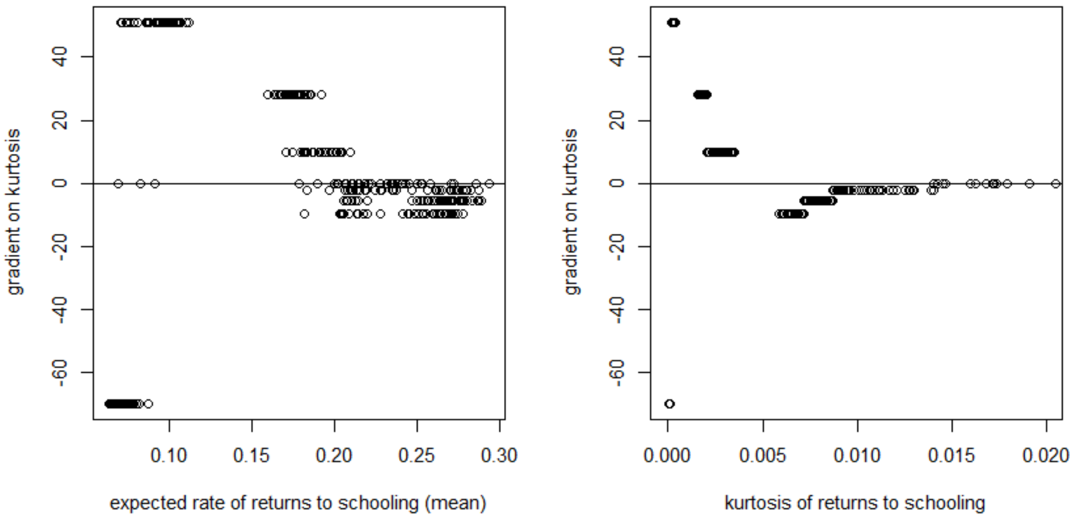


Figure 2.14: Statistically Significant Spline Gradients (Step 4) on 45° Plot

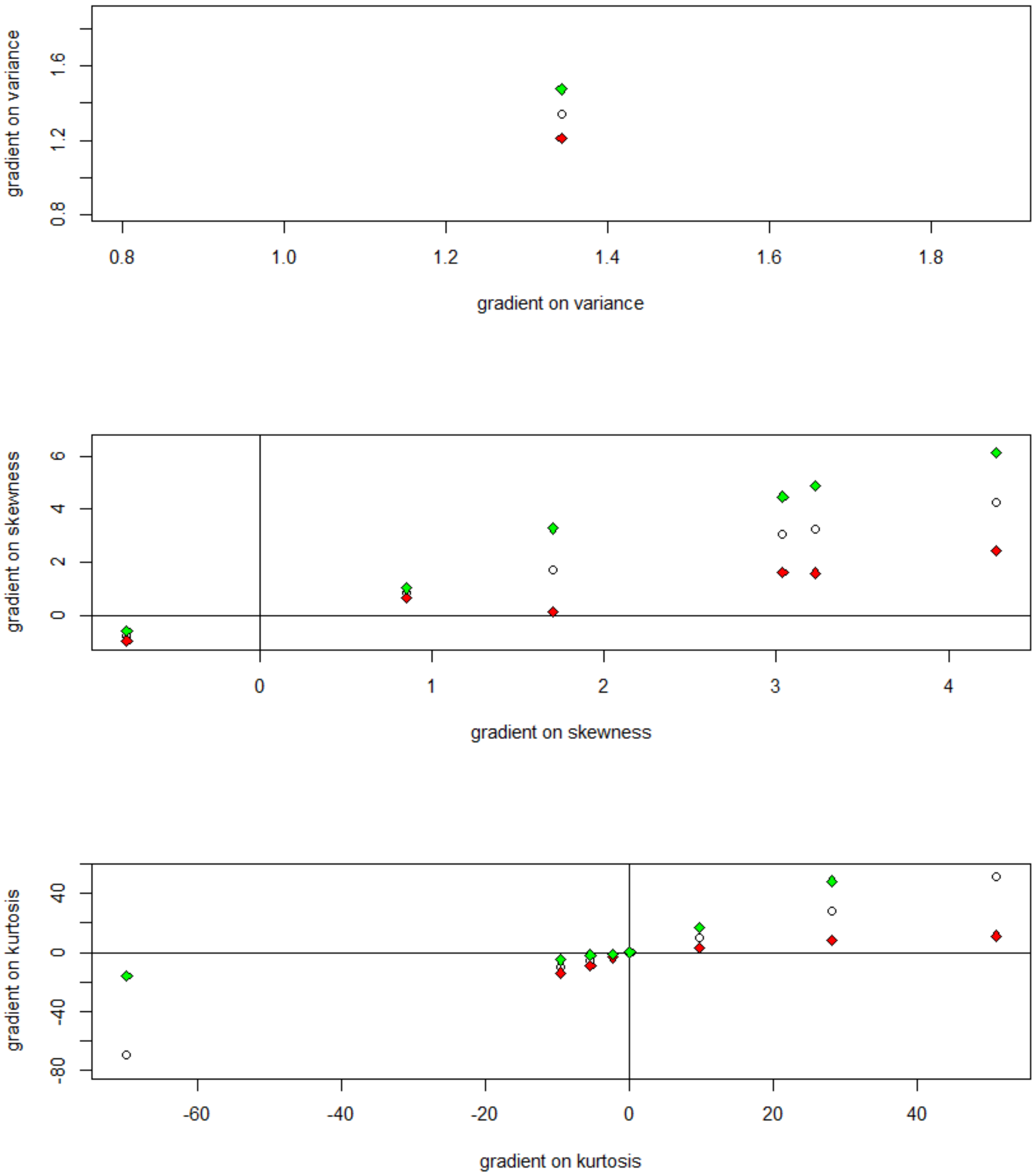
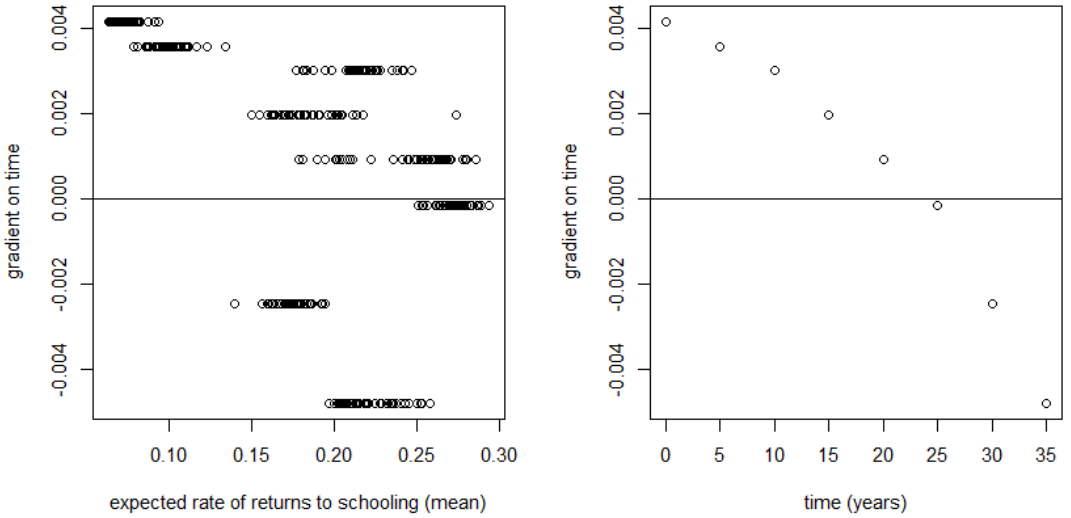


Figure 2.15: Spline Gradients on Time Trend from Step 4



all equal to 0.85. This means that, for example, decreasing the skewness from -0.01 to -0.02 (i.e. more risk) would result with a decrease in expected rate of returns to education (0.85*-0.01=-0.0082). However, there is only a few observations with negative skewness, so little can be concluded from this lack of variation.

We also observe that some of those distributions have a very high and statistically significant compensation for a slightly positively skewed distribution, while distributions with a slightly more positive skewness have a statistically significant negative compensation. The latter seems to indicate that individuals are willing to get paid less on average when they are in a more preferred positively skewed type of market. While the former suggests the opposite, that is, individuals require compensation for positive skewness as well. This means that they do not like either negatively or positively skewed returns distributions.

The effect of kurtosis on expected returns is nonlinear as well. There seem to be compensation for extremely light tails only, otherwise the effect is negative but close to zero. Though there seems to be a extremely high penalty and compensation for kurtosis that are equal to or very close to zero, this is probably due to the nature of our estimation method. It is fairly common for the tail observations to have extreme values in nonparametric estimations. The magnitude on slightly higher kurtosis are still relatively large but again this is not surprising considering how small the values of kurtosis are in our sample. Consider for example the few gradients on kurtosis that are equal to 9.81. If the kurtosis increased by 0.01 units (still a relatively big increase considering that most of the kurtosis values lie between 0 and 0.02) the expected returns would increase by 0.0981 units or 9.81 percentage points. Hence, those magnitudes are relatively important but not huge.

Figure 2.15 plots the gradients on time over time. The time trend effect seems to be overall very low. Interestingly, the expected returns seemed to have increased for the first 20 years in our sample (1980-2000), while it decreased on average for the last 10 years (2005-2015).

2.5 Conclusion

In this paper, we investigate whether education acts like an investment good. If so, parts of individuals' return to education could be a premium for the risk they incurred when investing in education. In other words, risk could play an important role in the educational decisions of risk-averse individuals.

To examine the existence of compensation for risk in education, we use the following steps. First, we obtain rate of returns to education for each individual in our dataset using categorical regression splines. We then split those rates of return to education by year and by state. From each subsample, we compute the first four moments: mean, variance, skewness, and kurtosis. Finally, we regress nonparametrically the mean of return to education on its higher moments. The gradients on those higher moments represent the compensation for each type of risk. Following the theoretical literature on decision making under uncertainty, we infer that risk-averse individuals favor small variance, positive skewness, and low kurtosis.

Overall, our results suggest that education acts like an investment good rather than a consumption good. We obtained statistically significant compensations for variance, which is in line with the previous empirical literature. Our results also show that there exists compensation for skewness and kurtosis. Those compensations are nonlinear with their respective moment, and are not always in the direction (gradients' sign) that theory predicts.

Specifically, the statistically significant gradients on skewness are mostly positive suggesting that individuals require a premium for positive skewness. This result goes against most of the empirical literature on educational investment which finds either a negative relationship between skewness and expected returns or no statically significant relationship.

While kurtosis has thus far not been studied in an educational investment context, our analysis finds that 80% of our gradients on kurtosis are statistically significant. Moreover, their magnitudes are not negligible. This suggests that kurtosis is also an important risk measure to consider when analyzing educational investment's behaviors.

Further analysis could help better explain the nonlinearities we observe. For example, sub-sampling our final stage further could show possible heterogeneity in risk and in risk compensation. The next chapter sub-samples the rates of return to education by occupation, by race, and by educational level.

CHAPTER 3
HOW EARNING COMPENSATION FOR RISK DIFFERS BY INDIVIDUAL CHARACTERISTICS? A SUBSAMPLE ANALYSIS

3.1 Introduction

Should I continue my education or should I start working? All students face this decision at every step in their educational path. Many factors will be taken into account when making this decision such as the cost of education, the potential wages of starting work now, and the potential wages from a higher degree. Those potential wages correspond to the returns of our educational investment and are uncertain. That is, as we are choosing between different educational options, we may be aware of the general shape of the distributions of returns to education for each option (i.e. risk) but we are unaware of where on that distribution we will fall (i.e. uncertainty). In other words, part of our decision making process when investing in education is taking into account the risk of the various potential path we could take and estimating their expected returns. If we do indeed take risk into account when making our educational decision, it means we see education as an investment good rather than a consumption good and as such we should observe trade-offs between expected returns to education and risks concerning those returns.

When we make this decision, we don't just choose how many years we will spend in school but also what career path (i.e. major, occupation,...) it will lead to. Not every year we spend in school is equal, even when controlling for all our personal observable characteristics. This sort of heterogeneity in return to schooling has already been observed in the literature. In this paper, we infer that not only the mean return to education faced by individuals varies across and within groups but also the risk. We build on the previous chapter to investigate further the heterogeneity in the risk and compensation for risk of educational investments within and between different groups. Specifically, we look at how risk and compensation for risk varies by occupations, by race, and also by education-level.

Using a nonparametric model will capture some potentially leftover heterogeneity, and we will be able to observe through sub-sample analysis whether there exists some variation in risk between different occupations. It is highly likely that some occupations are intrinsically more risky than others. However, it is less obvious whether those occupations are compensated for that higher risk with higher expected returns.

Risk and compensation for risk may also vary by race. When investing in education white individuals face very different risks (in terms of potential returns) than non-white individuals. That is, the shape of the rate of returns distribution might look very different for white individuals compared to non-white individuals. This effect could explain some of the racial disparity in expected rates of return to education. Or, on the contrary, we might see racial disparity in risk-return trade-offs i.e. that one group is not compensated as well as the other for incurring the same type of risk.

We also look at compensation by education level. If risk increases with schooling level and individuals are risk-averse, some of the rates of return to education differences might reflect compensation for risk. There might also be some time effect. If the rate of return to higher education increases faster than the rate of return to basic education, those with higher education (and initial higher earnings) will see their earnings go up more rapidly than those with lower levels of schooling (and lower initial earnings). This trend would worsen income distribution, other factors being equal.

3.2 Methodology

In this paper, we use the individual-level rates of return to education estimated in the previous chapter. That is, we estimated individuals' rate of return to education by regressing nonparametrically the log of individuals' weekly wage on education and other control variables for each of the 8 years in our sample[1](corresponds to step 1 and 2). However, contrary to the previous chapter which then split the sample by year and state to obtain year-state rates of return' distributions, here we split the sample in three different ways: by year and occupation, by region and race, and by region and education-level. Using those three types of distribution specification, we do three separate analysis where we compare the risk between and within each group and then regress the mean on the higher moments and observe whether trade-offs exist.

We split our full sample in three different ways:

1. by year t (8 years) and occupation j (24 occupations),
2. by year t (8 years), region j (9 regions), and race r (white vs. non-white),
3. by year t (8 years), region j (9 regions), and education level r (no high-school, high-school graduates, and college graduates).

That is, for each analysis separately we use either our obtained 192 year-occupation distributions, our 144 year-region-race distributions, or our 216 year-region-educ. distributions[2].

Then, for each 3 analysis separately, we obtain the first four (4) moments from each subsample:

$$Mean_{tj(r)} = r\bar{r}_{tj(r)};$$

$$Var_{tj(r)} = \frac{1}{N_{tj(r)}} \sum_i (rr_{itj(r)} - r\bar{r}_{tj(r)})^2;$$

$$Skew_{tj(r)} = \frac{1}{N_{tj(r)}} \sum_i (rr_{itj(r)} - r\bar{r}_{tj(r)})^3;$$

$$Kurt_{tj(r)} = \frac{1}{N_{tj}(r)} \sum_i (rr_{itj(r)} - r\bar{r}_{tj(r)})^4,$$

---

[1] Our sample, as described in the previous chapter, consist of working men between the age of 20 and 59 (working age) living in a household (excludes group quarters). For computational time reason, we only used the 8 quinquennial years between 1980 and 2015. The original data comes from the March Current Population Survey (CPS) dataset.

[2] Note that we used regions instead of states for the last two analysis because our sample size was not large enough to use states

where "Var" is the variance of each subsample, "Skew" is the skewness of each subsample (i.e. measures the degree and direction of asymmetry in the distribution), and "Kurt" is the kurtosis of each subsample (i.e. measures the heaviness of the tails).

Lastly, we use categorical regression spline (CRS) to regress the mean of returns to education on its corresponding higher moments (our measure of risk) for each subsample[3]:

$$Mean_{tj} = g(Var_{tj(r)}, Skew_{tj(r)}, Kurt_{t(r)j}) + u_{tj} \tag{3.1}$$

Similarly to the previous chapter, estimate the trade-offs nonparametrically because we suspect that the relationship between risk and return to education is nonlinear. As discussed then, by sub-sampling our analysis further we may capture some heterogeneity in risk-attitudes. For example, some occupations may be more prone than others to attract risk-seeking individuals, hence, sub-sampling by occupation may capture some of that effect. That is, occupations with more risk-averse individuals compared to other occupations may be better compensated. Also, some occupations are intrinsically "riskier" than others and the compensation for a generally low-risk occupation may be different than the compensation for a generally high-risk occupation.

Equivalent effects may be infer for the other two analysis. It is very likely that white men face very different risk in terms of rates of return distribution than non-white men. Generally speaking, white men may also have different risk attitudes than non-white men towards their education. Also, men with a college degree face very different rates of return distributions than men who did not complete high school.

3.3 Results

Using the estimated individual rates of returns to education obtained in the previous chapter "Risk-Return to Education Trade-Offs: To The Mean and Beyond!", we split them by year and successively: by occupation, by race (and region), and by education level (and region). From each group we get a distribution of rates of return to

---

[3] For all three analysis we also control for years using a time trend. Additionally, for each analysis respectively, we also control for: occupation fixed effects, region fixed effects and race (adding a white dummy), and region fixed effects and education level.

Table 3.1: Occupations Groups

Occ	**Occupation's Description**
1	Management, business, science, and arts
2	Business operations specialists
3	Financial specialists
4	Computer and mathematical
5	Architecture and engineering
6	Technicians
7	Life, physical, and social science
8	Legal
9	Education, training, and library
10	Arts, design, entertainment,sports, and media
11	Healthcare practioners and technical
12	Healthcare support
13	Protective service
14	Food preparation and serving
15	Building and grounds cleaning and maintenance
16	Personal care and service
17	Sales and related
18	Office and administrative support
19	Farming, fishing, and forestry
20	Construction
21	Extraction
22	Installation, maintenance, and repair
23	Production
24	Transportation and material moving

education and compute the first four moments: mean, variance, skewness, and kurtosis. We analyze how risk as measure by the higher moments varies between those different groups and over time. Then, we regress the mean (1st moment) of returns to education on their higher moments to see if we observe compensation for each different type of risk.

3.3.1 Analysis by Occupation

In this analysis, we split individual rates of return by year and by the 24 occupations[4] described in 3.1. We thus get 192 observations (=24 occupations times 8 years) for each moment. The summary statistics and the densities for each of those moments are presented in table 3.2 and figure 3.1 respectively. Note that, compared

---

[4] We use the occupation specifications provided by the CPS dataset. Note that occupations 25 and 26, which corresponds to the military and the unemployed respectively, are excluded from this analysis.

to our analysis by states (previous chapter) and if we ignore the extreme values, the distributions of expected returns and of the higher moments are very similar. That is, they have the same values for the median, 1st quartile, and 3rd quartile. The shapes of the distributions represented in 3.1 are also very similar. This suggest that there is the same amount of variation in risk between states as there are between occupations.

Table 3.2: Moments Summary Statistics (by occupation)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
mean	-0.04	0.14	0.20	0.19	0.25	0.34
variance	0.00	0.03	0.05	0.06	0.07	0.55
skewness	-2.55	0.00	0.01	0.00	0.02	0.54
kurtosis	0.00	0.00	0.00	0.08	0.01	11.99

Table 3.3 shows the correlation between each moment. The correlation between expected returns (mean) and all three higher moments is statistically significant. Contrary to our previous analysis by states, the correlation between the mean and the 4th moment kurtosis is negative and statistically significant. Also, the correlation with the variance is of much lower magnitude.

Table 3.3: Moments Correlation Matrix (by occupation)

	mean	variance	skewness	kurtosis
mean	1.00	0.43***	0.30***	-0.19***
variance	0.43***	1.00	-0.60***	0.75***
skewness	0.30***	-0.60***	1.00	-0.93***
kurtosis	-0.19***	0.75***	-0.93***	1.00
	Note:	*p<0.1;	**p<0.05;	***p<0.01

Table 3.4: Time Correlation Matrix (by occupation)

	mean	variance	skewness	kurtosis
year	0.59***	0.31***	0.14**	-0.10
	Note:	*p<0.1;	**p<0.05;	***p<0.01

Interestingly, risk within occupations as measured by variance and skewness seems to have increase over time. As described in table 3.4, variance is positively correlated with year at the 1% level while skewness is positively correlated with year at the 5% level. It is less obvious how kurtosis varied over time. The correlation between kurtosis and year is small, negative, and not statistically significant.

Figure 3.1: Moments Density (by occupation)

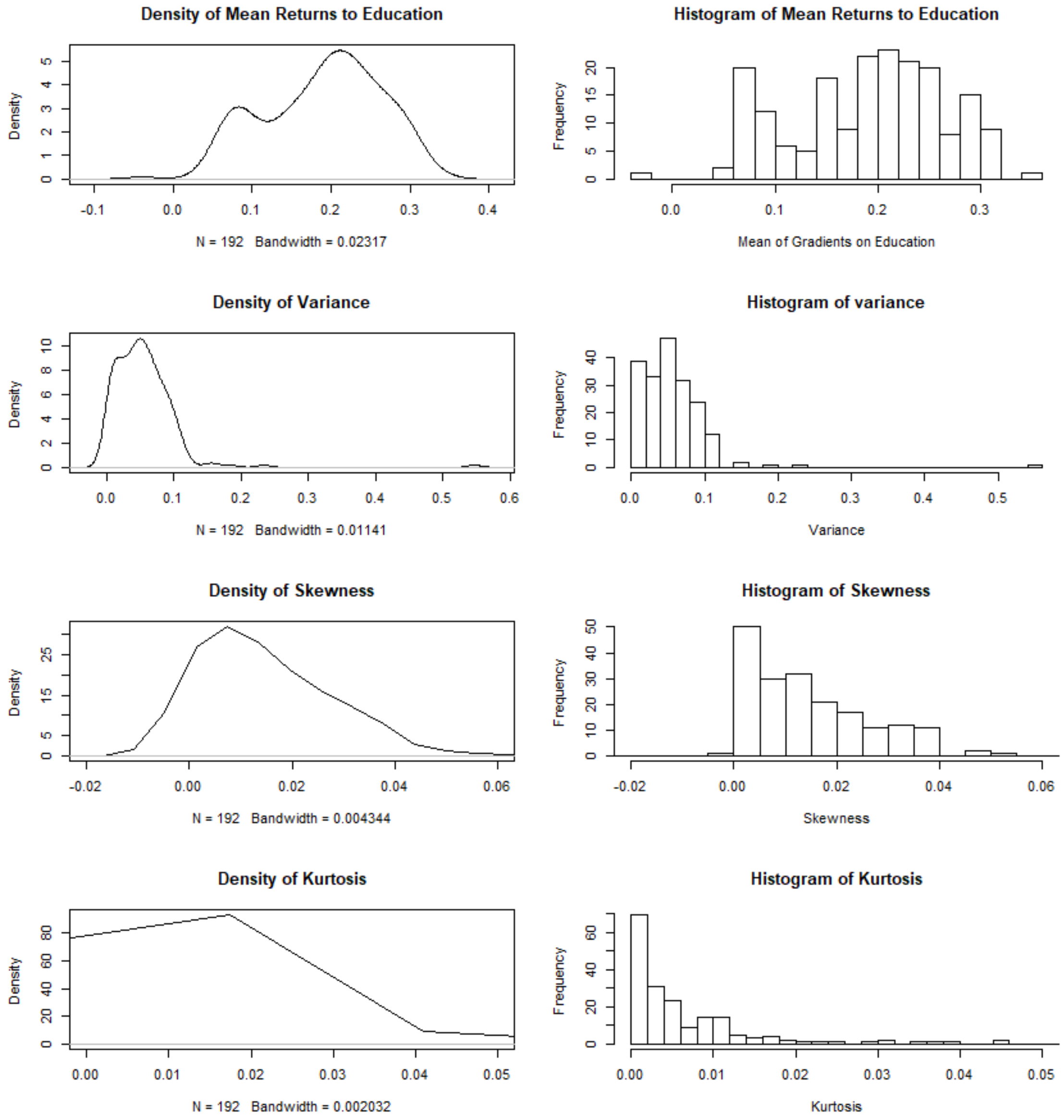
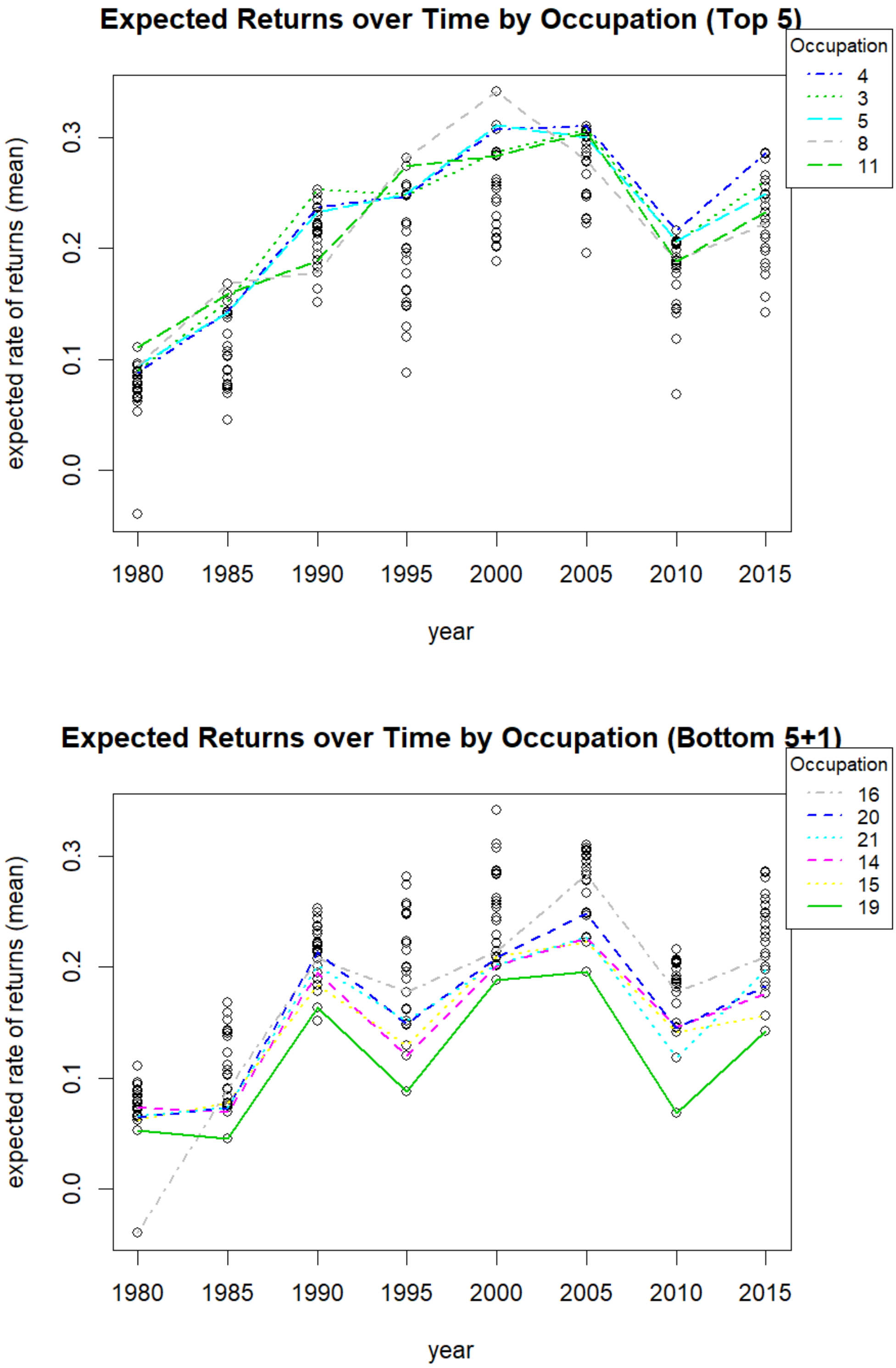


Figure 3.2: Expected Returns Top 5 and Bottom 5



Figures 3.2 - 3.5 plot each moment over time where each dot represents one occupation-year distribution. Each figure highlights the occupations with the highest values on average over the time period (top panel) and those with the lowest values on average over the period (bottom panel). For example, the occupation 4, i.e. computer and mathematical jobs, had on average the highest expected returns over the period (figure 3.2: top panel) while occupation 19, i.e. farming, fishing, and forestry jobs, had the lowest expected returns (figure 3.2: bottom panel).

Again, expected returns within occupations have been generally increasing over time as shown on figure 3.2. In the previous chapter, we noted that the expected returns by state have been decreasing both between 1990 and 1995 and between 2005 and 2010. While we see here the same clear effect between 2005 and 2010 where every occupations saw their expected return decrease, the effect between 1990 and 1995 is more heterogeneous. The former effect is unsurprising considering the wide effect of the 2008 financial crisis. It seems though, that the cause of the overall decrease in states' expected returns (probably the small recession of 1990) did not affect all occupations the same way. While some occupations saw their expected returns decrease, others saw it flatten or even increase. Specifically, between 1990 and 1995 occupations 8 and 9 (legal and education related jobs) strongly increased while occupations 4, 5, 7, and 11 slightly increased. All other occupations either flattened or decreased between 1990 and 1995. The five biggest decrease between those two years were in healthcare support jobs, farming, fishing, and forestry jobs, food preparation and serving jobs, production jobs, and transportation jobs. The last two occupations are to be expected considering how the 1990s recession drove up the world price of oil, while the others might have been more affected than others by the early 90's decreased in consumer confidence.

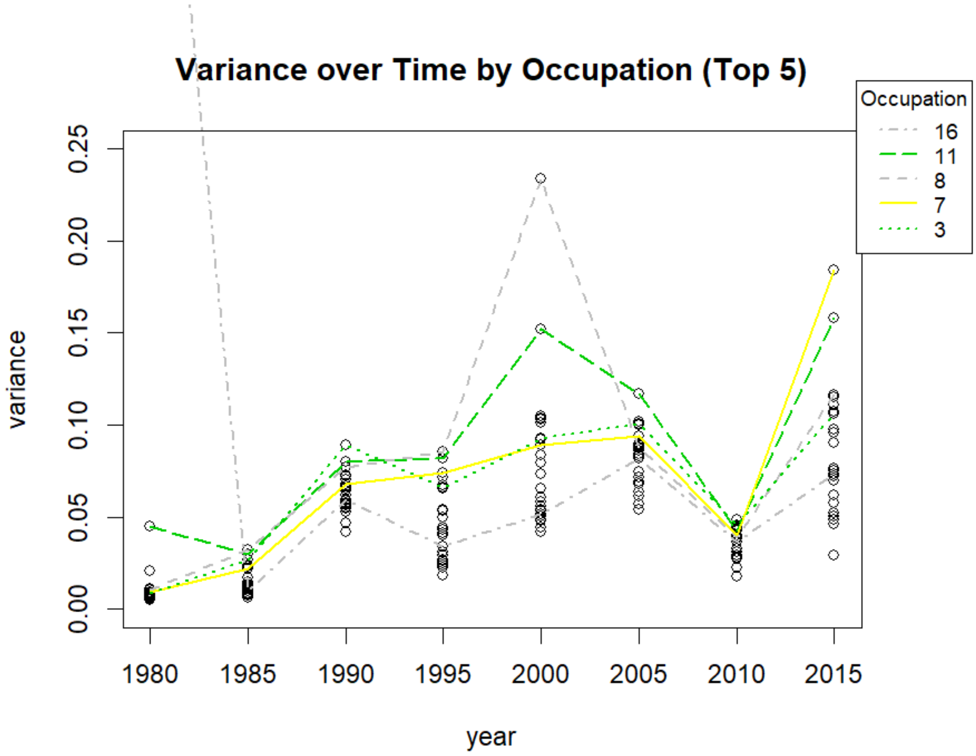
Note that in the bottom panel of figure 3.2 I added occupation 16, i.e. personal care and service jobs, to the bottom 5 because of it's unusually low expected return in 1980. It is, however, not the 6th lowest average expected return. This occupation also had in 1980 the highest variance, lowest skewness, and the highest kurtosis in the whole sample. In other words, choosing to work in a personal care and service type job in 1980 was not only very risky but yield very low (negative) returns. This occupation was, however, much closer to average in terms of both expected returns and risk all of the other years.

Figure 3.3 shows a very similar increase in the variance over time as the variance by state (previous chapter). Interestingly, occupations that tend to have low variance also tends to have low expected return and vice versa (with the exception of personal care jobs). As you can see comparing figures 3.2 and 3.3, some of the occupations in the top and bottom of expected returns are also in the top and bottom of variance.

Note that occupation 11, i.e. healthcare practitioners, tends to have consistently high variance throughout time (always being in the top 2) suggesting that this is particularly risky occupation. Occupation 8, i.e. legal jobs, is also often at the top in terms of variance. Those two occupations also tend to have high skewness and high kurtosis. They especially stand out in 2000 with unusually high variance, skewness and kurtosis. This suggest very high inequality in terms of returns to on more year of education within those two occupations. That is, no only were there a lot of variation in returns to education within healthcare practitioners and legal jobs but there were also a few individuals with extremely high rates of returns while most were on the lower side of returns. The high kurtosis suggest that those individuals had both a higher chance to be close to the expected return but also a higher chance to be extremely far away (i.e. a lower chance to have returns between the mean and the extreme high and low returns). This variation, however, seems to have been compensate by higher expected returns compared to other occupations (see figure 3.2).

Low paying jobs like farming, fishing, and forestry jobs (19) or extraction jobs (21) not only tend to have low variance, skewness, and kurtosis, but they also see less variation in those measures from year to year. In other words, they are consistently low risk in the sense that everyone in those occupations tend to have about the same rate of returns to education. However, the expected return in those occupations seems to vary a lot more from year to year than other occupations. This difference compared to other type of jobs is mainly driven by the the effect of the early 1990s recession which

Figure 3.3: Variance Top 5 and Bottom 5



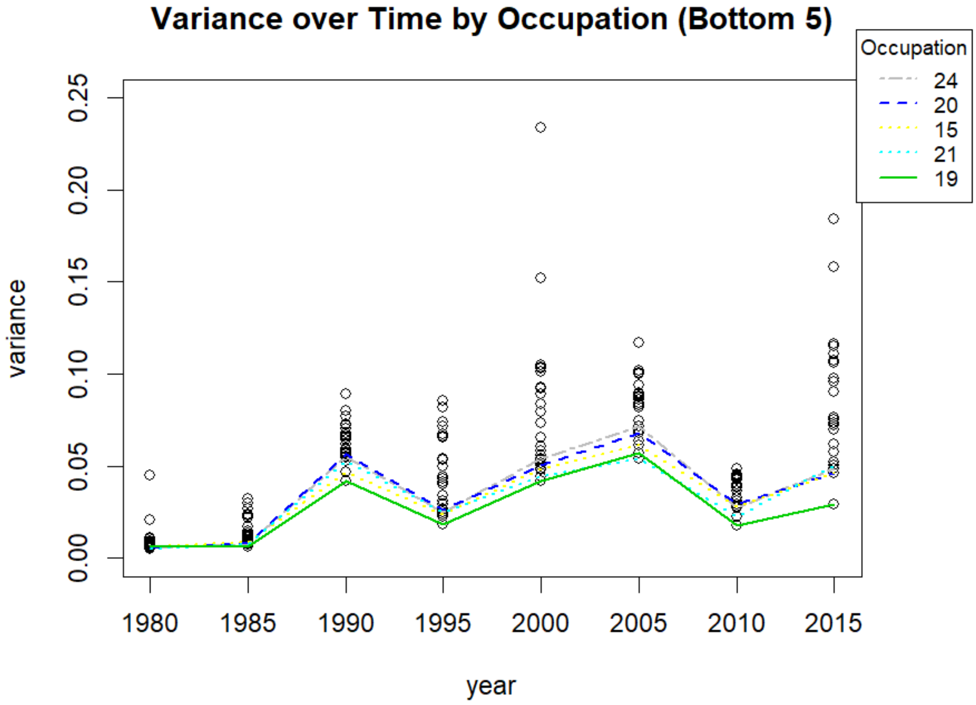
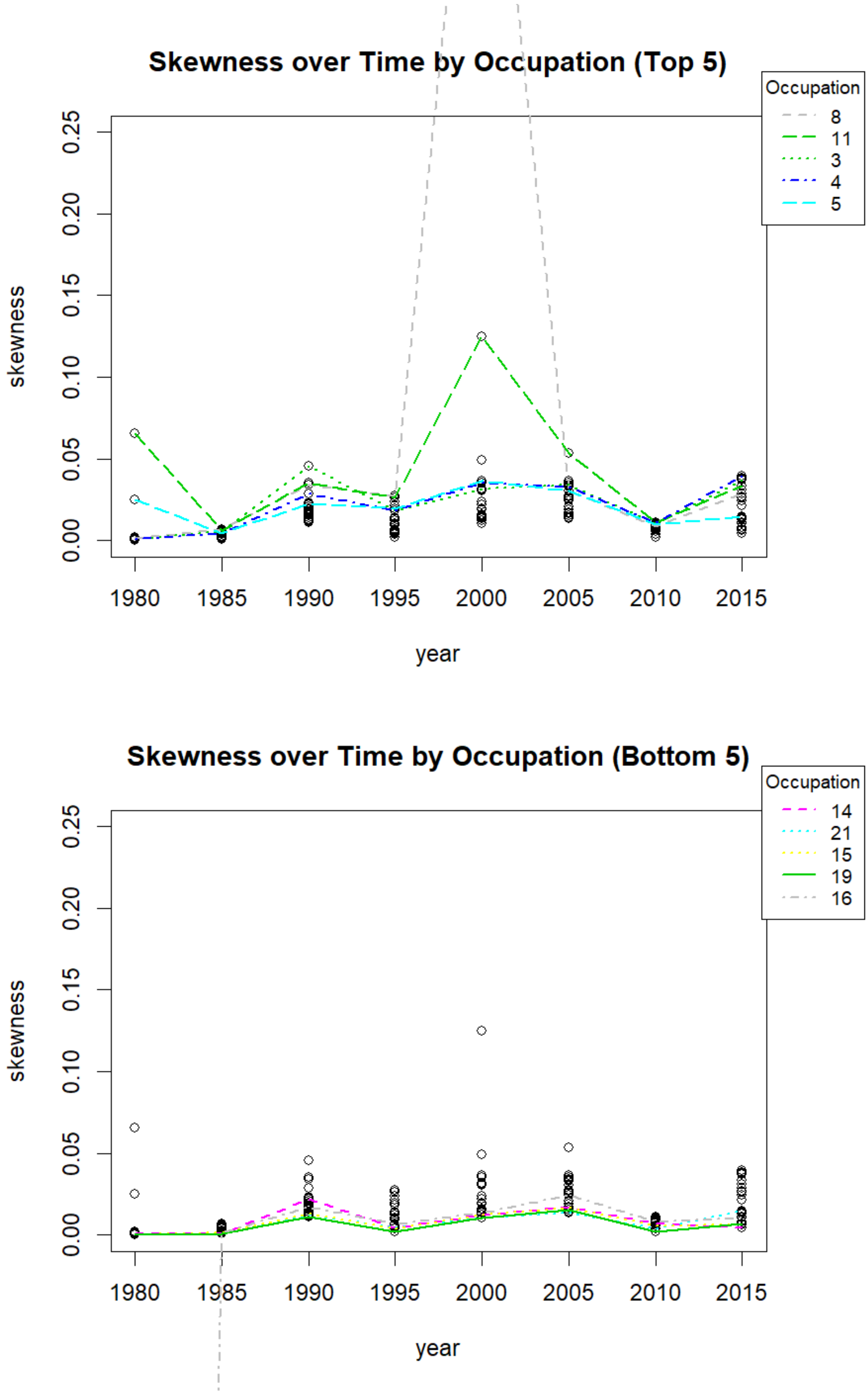


Figure 3.4: Skewness Top 5 and Bottom 5



seems to have affect those jobs more than other jobs that are maybe less affected by oil markets.

Figures 3.4 and 3.5 show skewness and kurtosis over time respectively. Similarly to the by state analysis (previous chapter), overall both measures do not show much variation from year to year. The years 1980, 2000, and 2015, however, display a few occupations with extreme or higher than usual values. This suggests that the risk faced by individuals during those 3 years was particularity unequal between occupations. Oddly, our previous by state analysis showed that the outliers in 2000 were Minnesota and Kansas, whereas here, the outliers are legal jobs and healthcare practitioners (8 and 11). There are, however, no particularly large amount of men working in legal and healthcare type jobs in those two states. The most common jobs is Minnesota and Kansas over the period are farmers and truck drivers[5].

Interestingly, the 2008 crisis seems to have both decrease the risk between (lower range) and within (lower values) occupations (see variance, skewness, and kurtosis in 2005 vs. 2010). However, the risk in 2015 seems to be worse than in 2005 both within and between occupations suggesting there were more inequality in terms of returns to education a few years after the 2008 recession ended than there were a few years before it began.

Table 3.5: Spline Gradients Summary Statistics (by occupation)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
g.variance	-1.37	2.10	3.51	3.14	4.13	26.13
g.skewness	0.35	0.64	0.74	0.71	0.77	0.96
g.kurtosis	-23.44	-9.21	-5.66	-3.19	-0.81	38.57

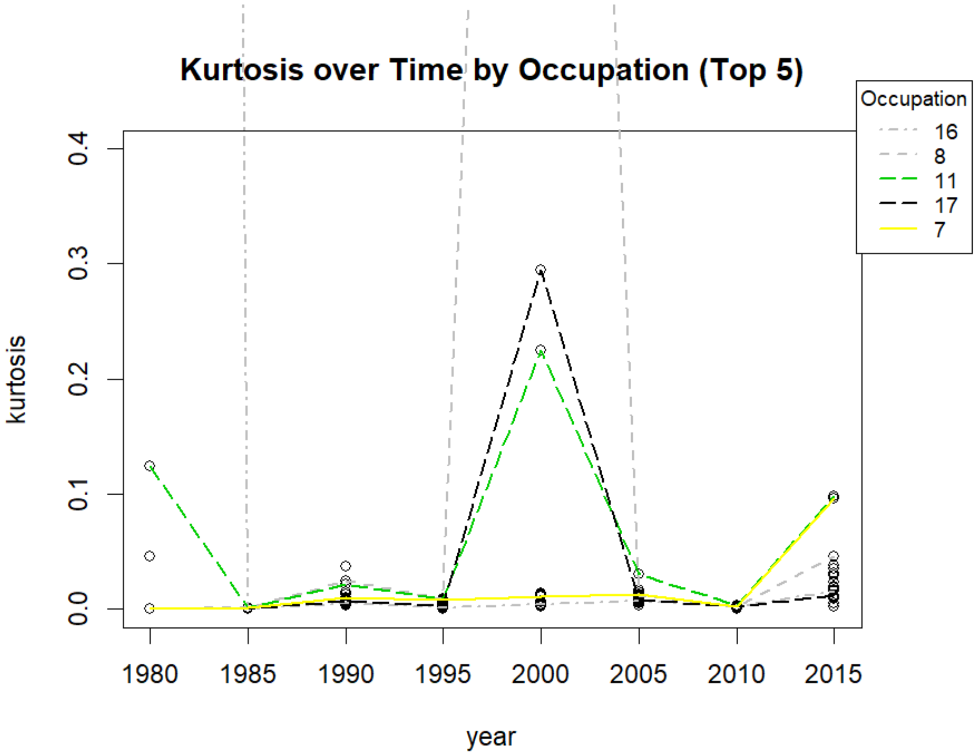
Table 3.6: Spline Gradients Summary Statistics (by occupation): Statistically Significant Results Only

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	obs.
g.variance	-1.37	2.12	3.46	3.05	4.15	5.30	179
g.skewness	0.62	0.75	0.75	0.75	0.77	0.86	96
g.kurtosis	-19.19	-11.43	-5.75	-8.08	-5.50	-0.09	102

---

[5] https://www.npr.org/sections/money/2015/02/05/382664837/map-the-most-common-job-in-every-state

Figure 3.5: Kurtosis Top 5 and Bottom 5



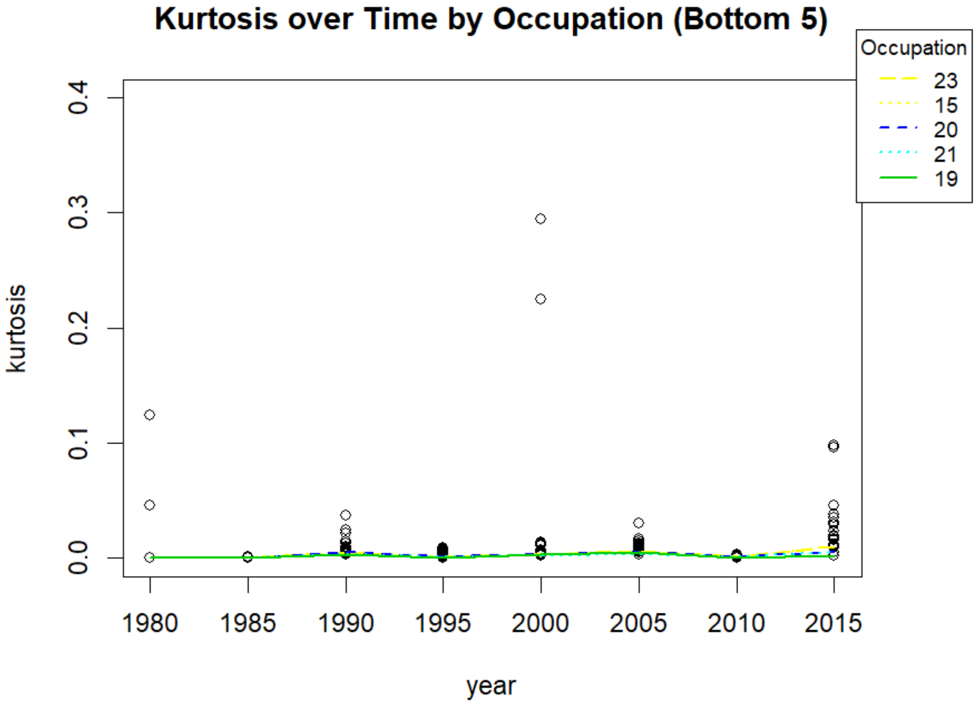


Table 3.7: OLS Results (by occupation)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.10	0.01	7.68	0.00
variance	1.75	0.11	15.51	0.00
skewness	0.03	0.04	0.64	0.52
kurtosis	-0.09	0.01	-7.63	0.00
time	0.00	0.00	2.52	0.01

The results from regressing the mean on the higher moments and controlling for time and occupations fixed effects can be found in tables 3.5 - 3.7. Tables 3.5 and 3.6 show the results using categorical regression splines (CRS) while table 3.7 shows the results using OLS. For a more visual representation of the CRS gradients, figure 3.6 show the gradients' density and histogram on each moment respectively. Compared to CRS, OLS seems to underestimate the effect of all three types of risk on expected returns. That is, all of the OLS coefficients on the higher moments are closer to zero than their respective 1st quartile value of statistically significant gradients (Table 3.6).

OLS find that skewness has no effect on expected returns. This would suggest that occupations with high skewness, i.e. where most individuals are on the low end of the distribution and a few individuals have very high returns, are not compensated by higher expected returns (and are not hurt through lower expected returns either). However, the CRS results suggest that while skewness does not matter for half of our sample, it does matter for the other half (96 out for 192 are statistically significant[6]) The gradients on skewness are found to be much lower magnitude than in our previous by state analysis. While the OLS coefficients are found to be statistically significant on both variance and kurtosis, only slightly more than half of the CRS gradients on kurtosis are found to be statistically significant.

Specifically, none of the gradients on skewness are statistically significant for all occupations in 1985 and 2010 (see figures 3.7 and 3.8). This is not surprising considering that skewness during those two years were overall very close together (between occupations) and of small values (see figure 3.4). Only occupation 5 and 11 have sta-

---

[6] As described in the previous chapter, we use the 95% confidence bounds of our estimates to define statistical significance.

Figure 3.6: CRS Gradients Density and Histogram

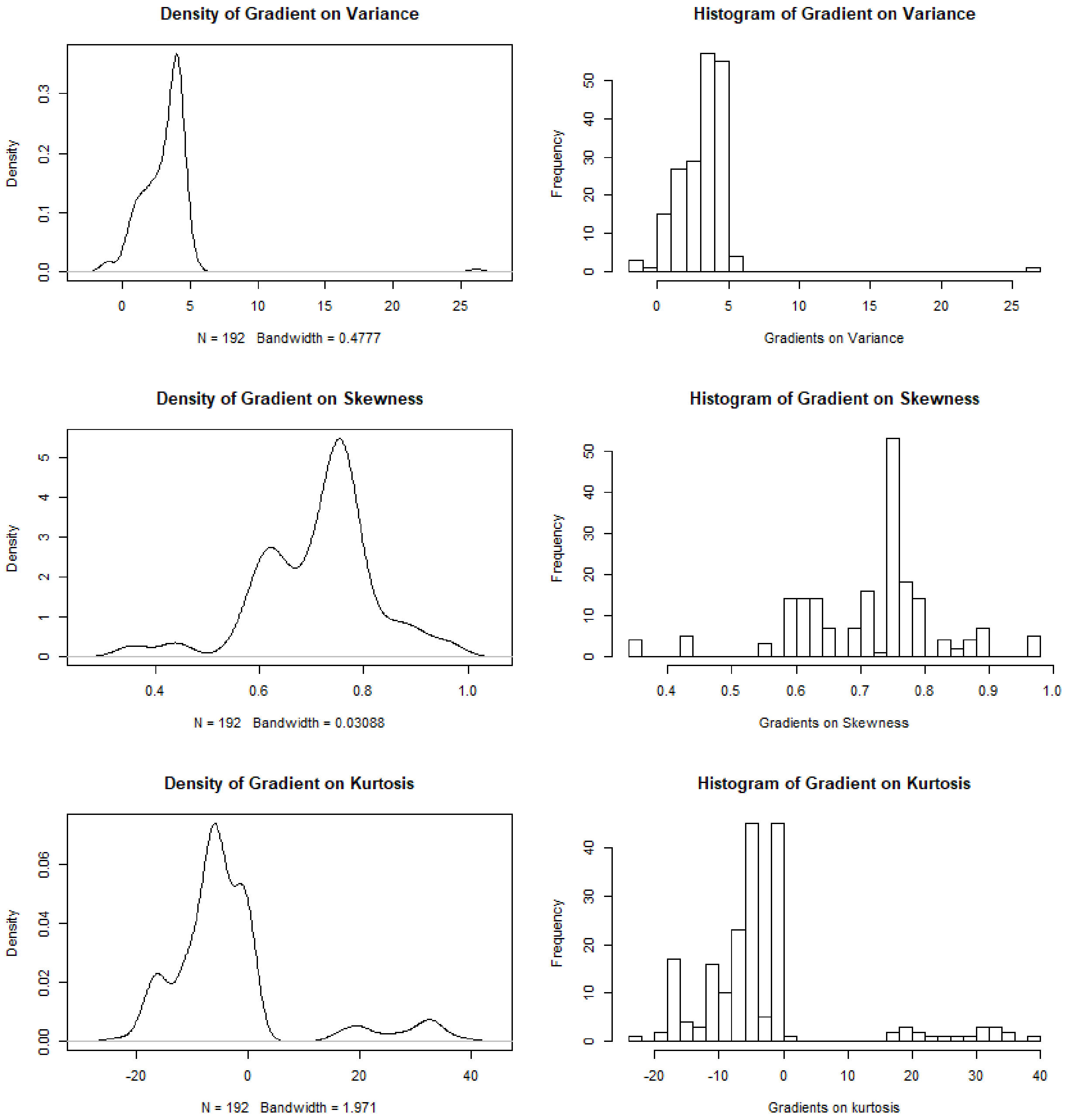


Figure 3.7: CRS Gradients over Time

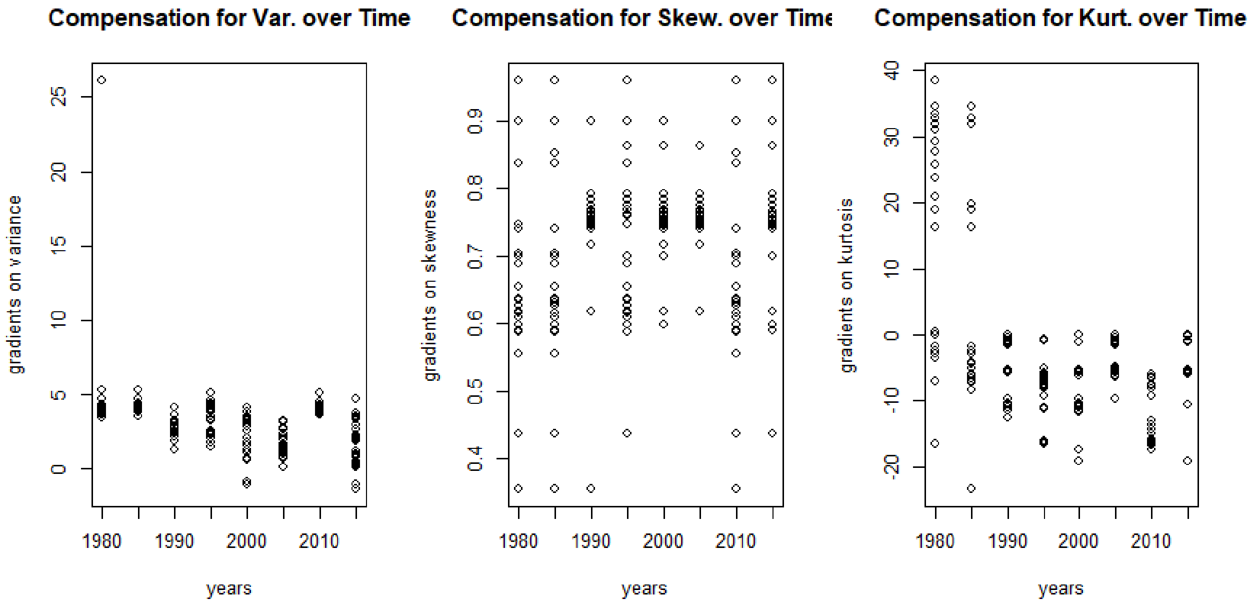
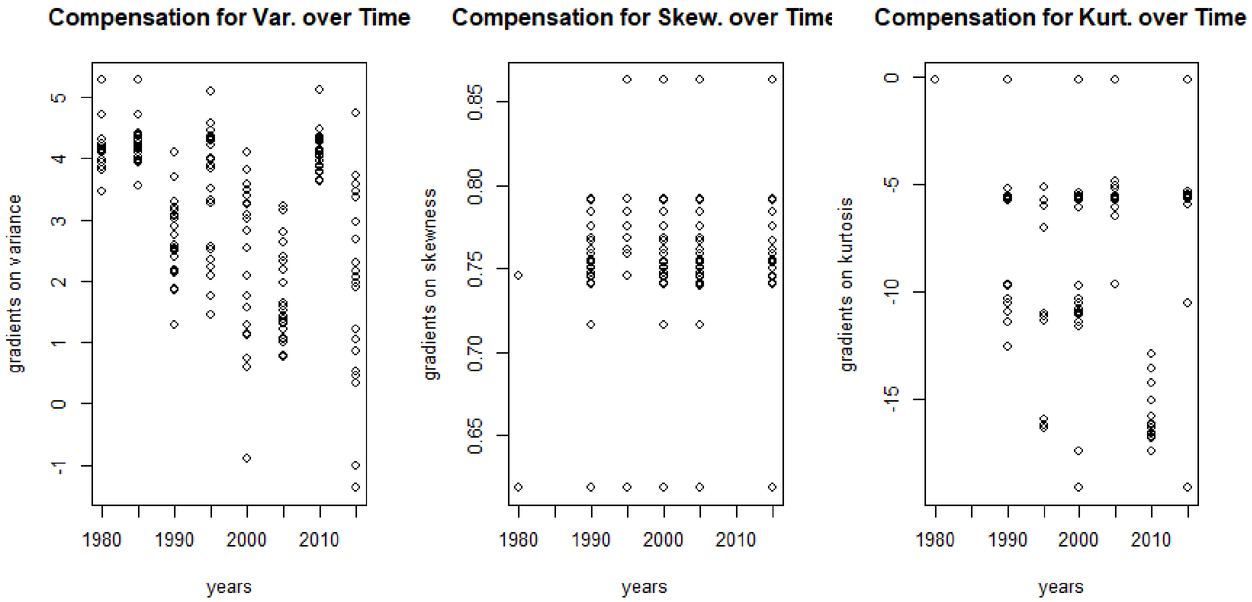


Figure 3.8: CRS Gradients over Time (Statistically Significant Only)



tistically significant gradients on skewness in 1980. Again, this results is not surprising as those two occupations had unusually high skewness in 1980 compared to other occupations. The other year that stands out is 1995 with 9 occupations with gradients on skewness that are not statistically significant. Looking at figure 3.4, we can see that 1995 has slightly lower skewness values overall compared to 1990, 2000, 2005, and 2015, with occupations on the lower bound that have skewness equal or extremely close to zero. Also note, that the gradients on skewness tend to stay the same over time for each occupation separately.

A very similar story can explain which occupations and years have statistically significant gradients on kurtosis. Note on figure 3.8 that none of the positive gradients from 1980 and 1985 were statistically significant. Surprisingly, though all of the kurtosis in 2010 were very close to zero, a lot of the gradients (on kurtosis) in 2010 are statistically significant and with high negative magnitude. Also, contrary to the gradients on skewness, the gradients on kurtosis vary a lot within a same occupation over time.

Interestingly, the gradients on skewness describe a clear positive relationship with expected returns while the gradients on kurtosis show a clear negative relationship with expected returns. This goes against the theoretical predictions of decision making under uncertainty which expect individuals to be prudent and seek compensation for negative skewness (not positive skewness) and to be temperate and seek compensation for higher kurtosis. This contrary effect is similar to what we found in the previous chapter (by state analysis).

Figures 3.9 - 3.11 show how the gradients on each moment vary with expected returns and with their respective moment. Overall, both compensation for variance and for kurtosis seem to vary with expected returns and with their respective moment following a particular trend. On the other hand, the gradients on skewness do no seem to vary in an explicit manner.

As shown on figure 3.9, compensation for variance seems to be high for low level of expected return and for low level of variance. As variance increases though, the compensation for variance decreases. Meaning, as the risk measured by variance increases,

Figure 3.9: Spline Gradients on Variance (Stat. Sign. Only)

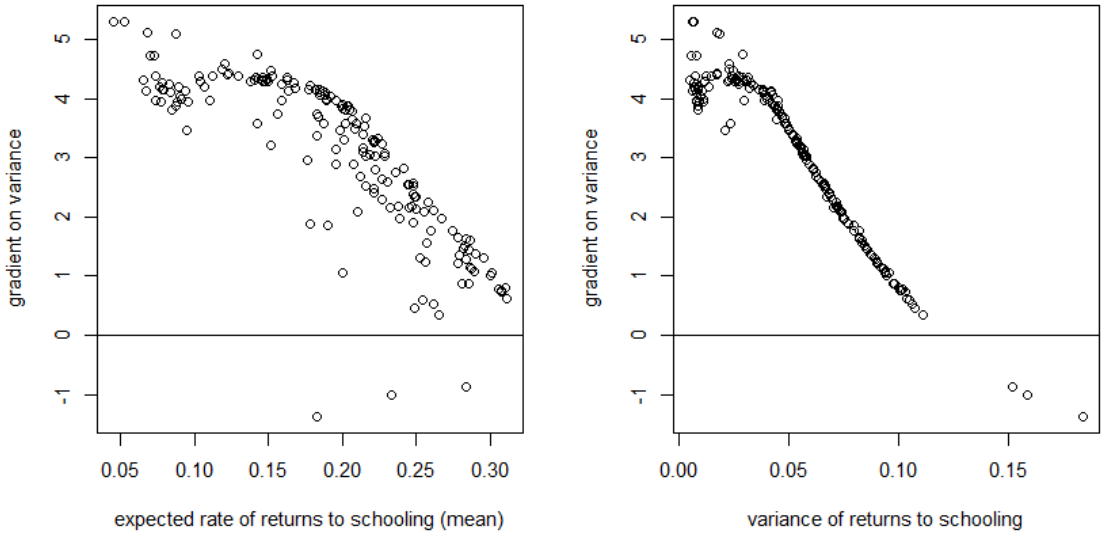


Figure 3.10: Spline Gradients on Skewness (Stat. Sign. Only)

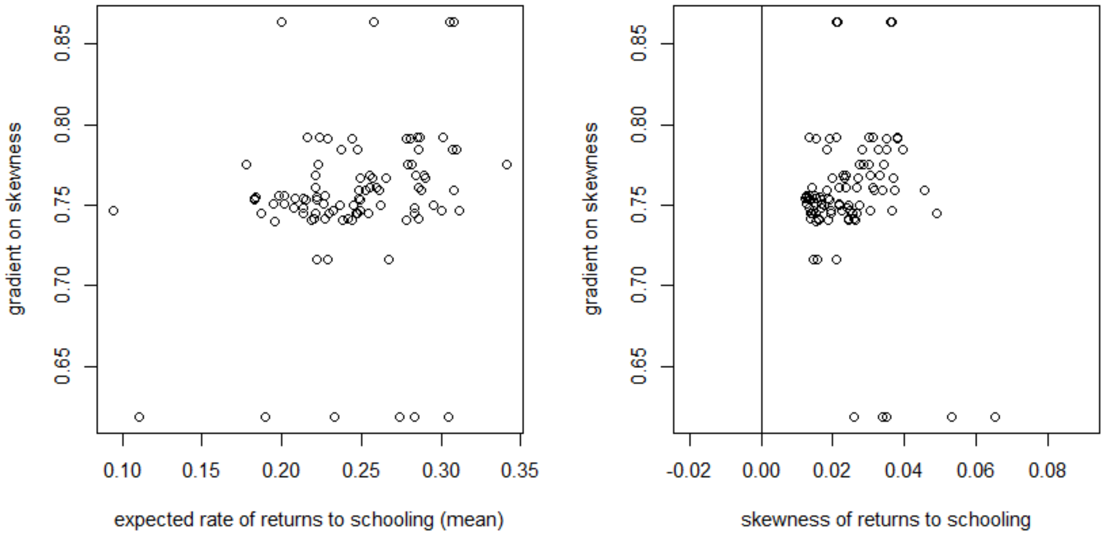
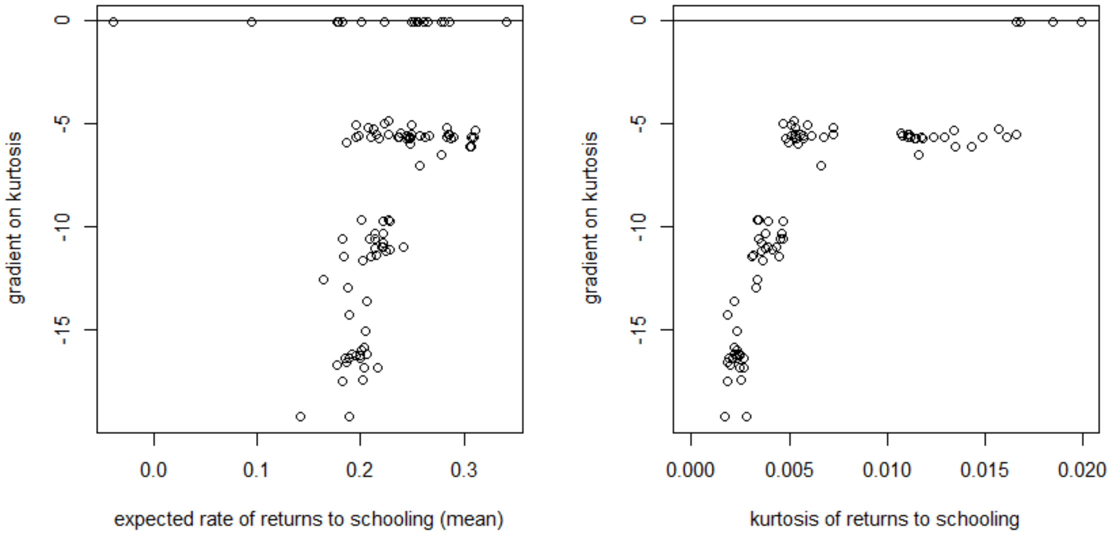


Figure 3.11: Spline Gradients on Kurtosis (Stat. Sign. Only)



individuals are still compensated with higher mean returns (gradients are still positive) but to a lesser extend (lower magnitude). For example, at a variance of 0.04 the gradients are around 4. That is, at this level of variance (0.04), increasing variance by 0.01 units will be compensated by an increase in expected return of about 0.04 units on average (or 4 percentage points). On the other hand, if the variance is higher at about 0.10 units and it the increase by 0.01 units, expected return will increase by about 0.005 on average (or 0.5 percentage points).

The gradients of skewness doe no seem to have any obvious relationship with expected returns and with skewness itself (figure 3.10). The range of the gradients is relatively small. That is when skewness has any effect, increasing skewness by 0.01 units will be compensated by an increase in expected return of at least 0.0062 units and at most 0.0085 units (or 0.62 percentage points and 0.86 percentage points respectively).

Both expected returns to education and kurtosis seem to have generally a positively sloped relationship with the gradients on kurtosis. As shown by the right panel of figure 3.11, the gradients on kurtosis become less negative (close to zero) as kurtosis increases for kurtosis less than 0.006. However, when kurtosis is higher than 0.006 units, the gradients on kurtosis are all around -5. As most kurtosis are within 0 and

0.02 units, let's consider an increase in kurtosis of 0.001 units. If kurtosis is already higher than 0.006 and increase further by 0.001 units, expected return will decrease by 0.005 units on average (or 0.5 percentage points). If kurtosis is low at about 0.003 and the gradient is about -17, then increasing kurtosis by 0.001 units will result in a decrease in expected returns of about 0.017 units on average (or about 1.7 percentage points).

3.3.2 Analysis by Race

The analysis by race divides the individuals' rate of return to education by region and by race (white vs. non-white). Table 3.8 describes which states are included in each of the 9 regions used. Note that we also did the analysis be race only to get a national average of the difference between whites and non-whites.

Table 3.8: Regions Description

Region	**Region Name**	**States Included**
11	New England Division	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
12	Middle Atlantic Division	New Jersey, New York, Pennsylvania
21	East North Central Division	Illinois, Indiana, Michigan, Ohio, Wisconsin
22	West North Central Division	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
31	South Atlantic Division	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
32	East South Central Division	Alabama, Kentucky, Mississippi, Tennessee
33	West South Central Division	Arkansas, Louisiana , Oklahoma, Texas
41	Mountain Division	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
42	Pacific Division	Alaska, California, Hawaii, Oregon, Washington

While figure 3.12 shows the difference in expected rate of return to education

Figure 3.12: Expected Returns by Race (nationally)

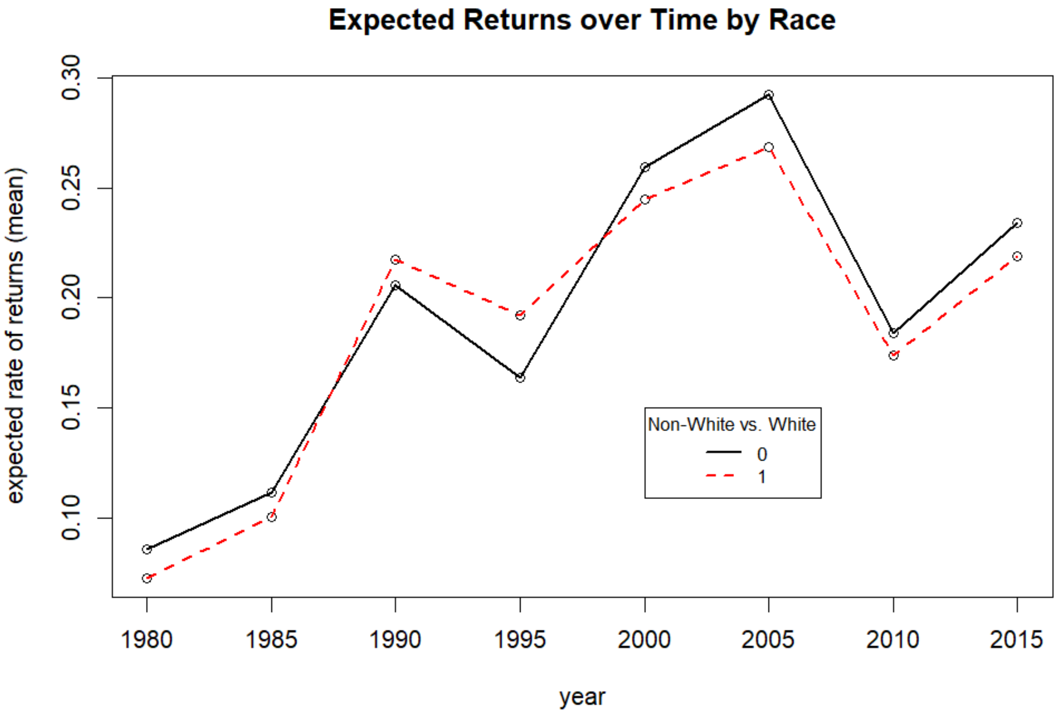
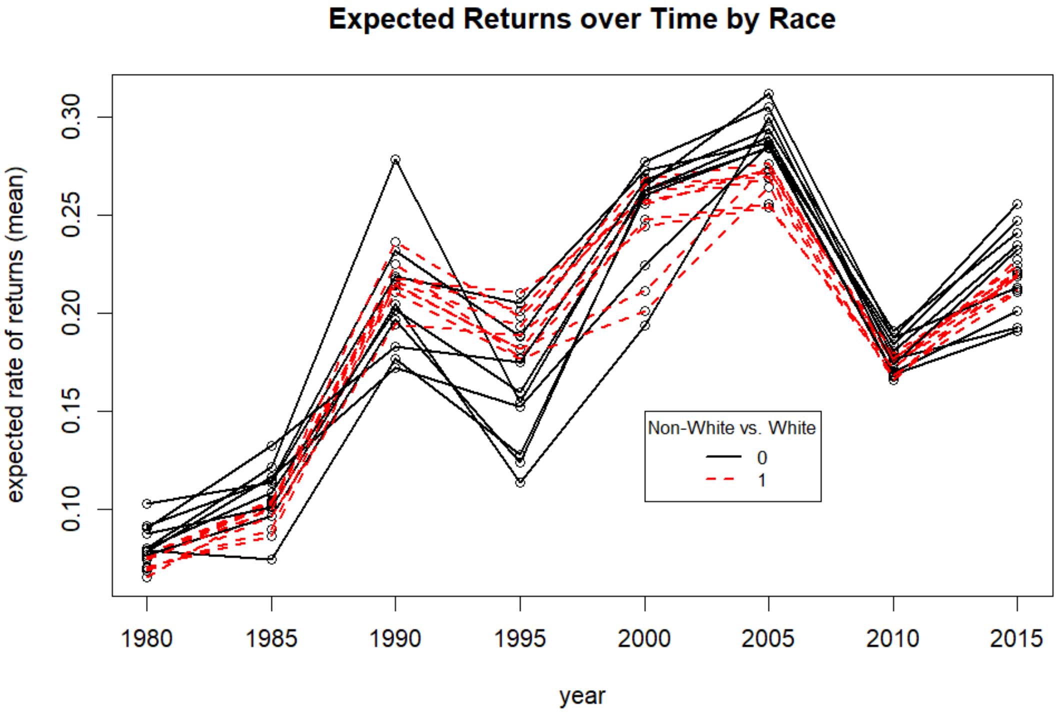


Figure 3.13: Expected Returns by Race (by region)



between whites and non-whites nationally, figure 3.13 shows the same differences but on a regional level. Nationally, apart from years 1990 and 1995, expected rates of return for non-whites (solid black line) have been higher than for whites (dashed red line) over time. Interestingly, looking at figure 3.13, it seems that overall the expected rates of return of non-whites (solid black lines) vary a lot more by regions than those of whites (dashed red lines) with the exception of 2000.

Overall, the regional graphs looked very similar to the national graph shown in figure 3.12. One notable difference is in 2015 where the expected rates of return for whites was higher than for non-whites in the New England Division, the East North Central Division, the West North Central Division, and the East South Central Division. Specifically, comparing the two close southern regions 31 and 32 (figures 3.14 and 3.15), we see that they have very similar patterns over the whole period with the exception of 2015 where non-whites' expected rates of return go in completely opposite direction. That is, in 2015, non-whites in the South Atlantic Divsion have the highest expected rate of return between all regions while non-whites in the East South Central Division have the lowest expected rate of return to education.

On a national level, risk as measured by variance seems to be similar between whites and non-whites up until 2000 (see figure 3.16). In 2005, a disparity between whites and non-whites starts to form which diminishes right after the 2008 financial crisis but exacerbates a few years later. Non-whites in 2015 end up facing much higher variance in 2015 than whites. This indicates that there is a lot more variation in rates of return between non-whites than between whites in 2015. When separating the distributions by regions as well (figure 3.17), we note that all of the distributions for non-whites have higher variance than those for whites in 2005, 2010, and 2015. Note that the top three highest variance for non-whites in 2015 are in the Middle Atlantic Division, the New England Division, and the East North Central Division respectively. On the other hand, the West North Central Division had almost no gap between the expected rate of return of whites and of non-whites.

Though there is almost no gap between whites' and non-whites' variance for the

Figure 3.14: Expected Returns by Race (South Atlantic Division)

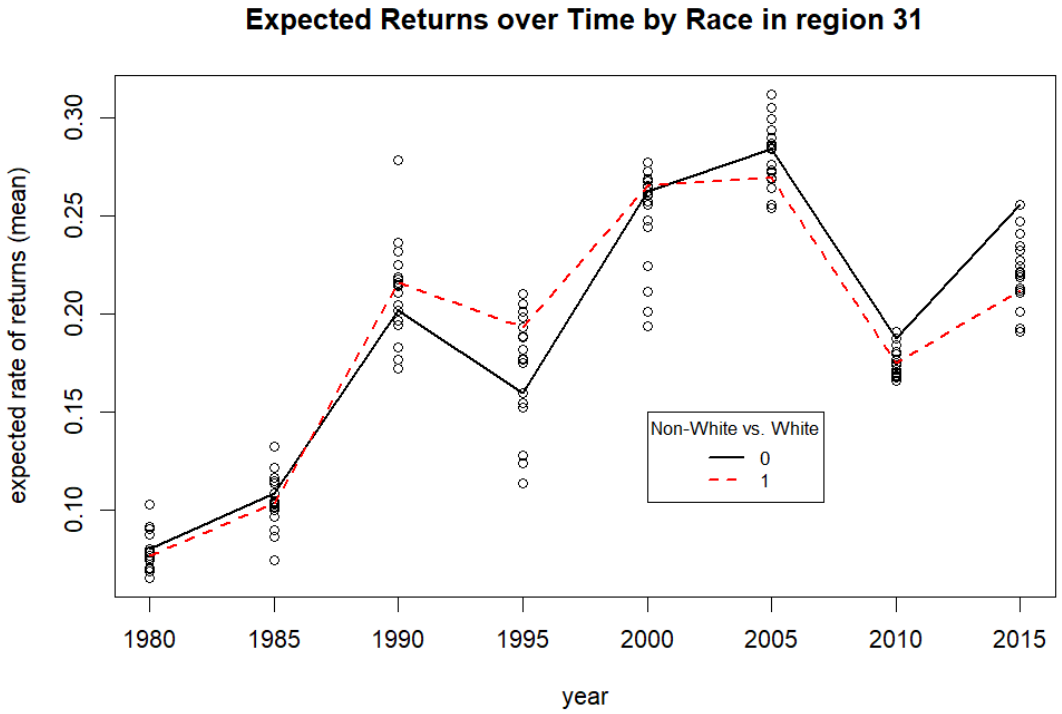


Figure 3.15: Expected Returns by Race (East South Central Division)

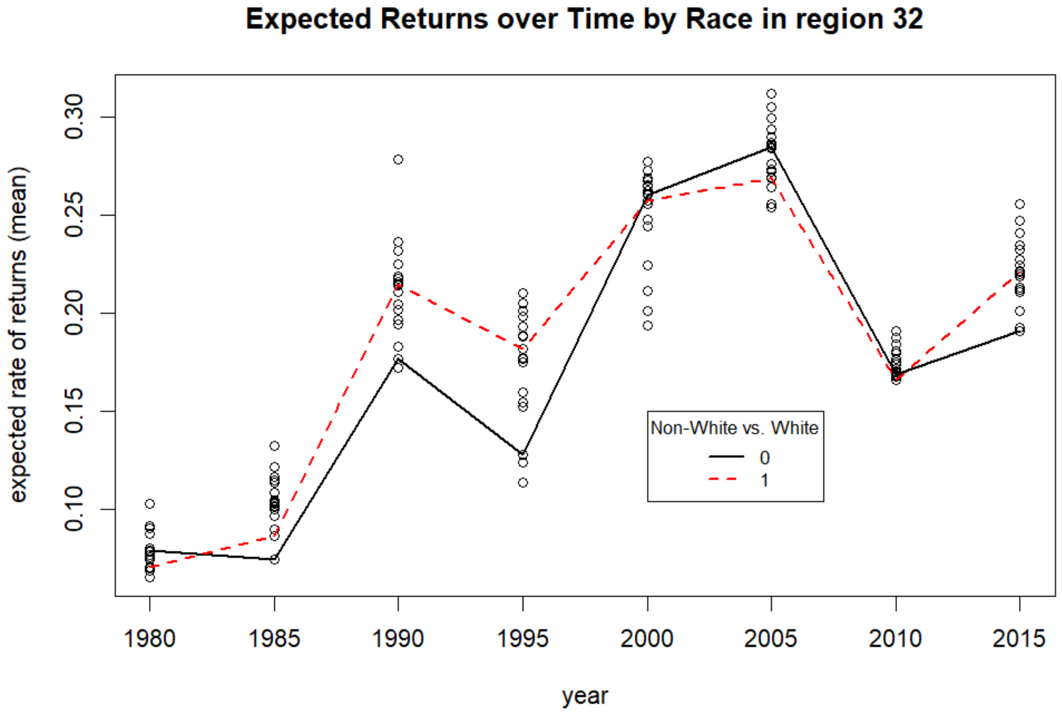


Figure 3.16: Variance by Race (nationally)

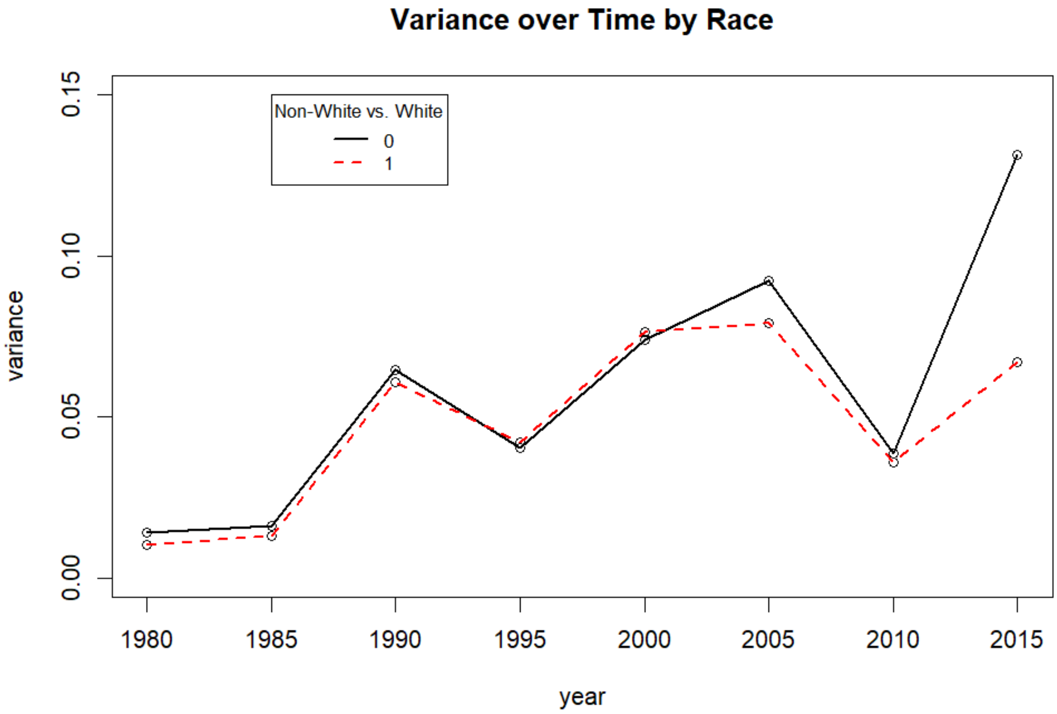
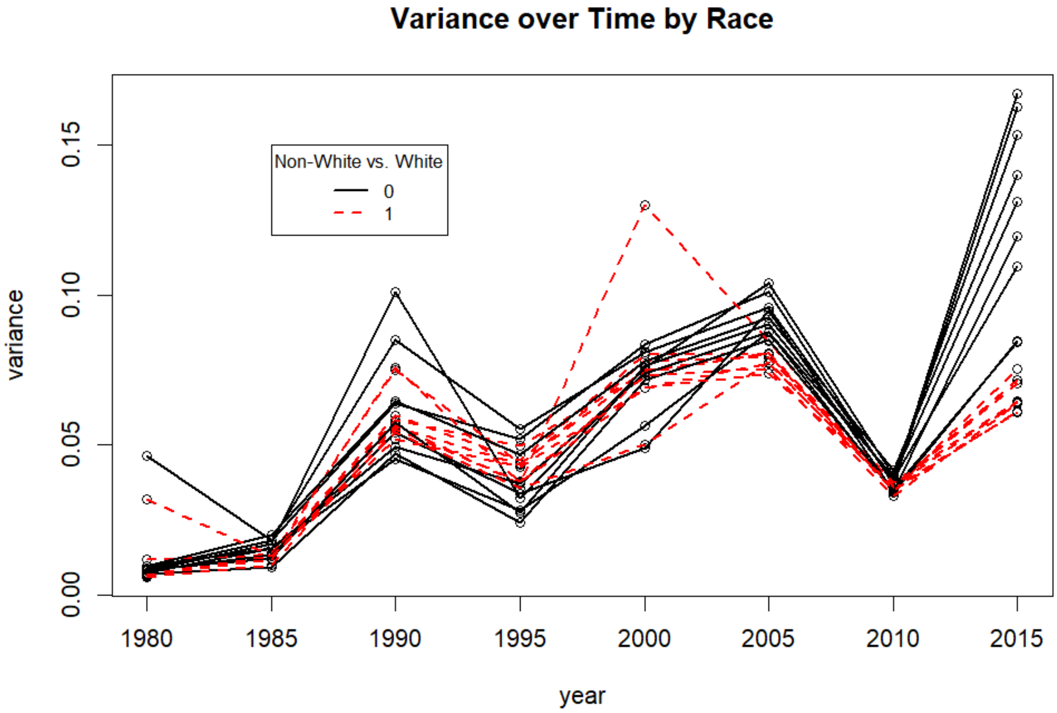


Figure 3.17: Variance by Race (by region)



whole 1980-2000 period on a national level, there exist some gaps when looking at each region separately. While region 41 looks very similar to the national graph on figure 3.16, the rest of the regions fall into two different camps. One camp has non-whites' variance being higher than whites' variance for most of or the whole 1980-2015 period. This camp is mainly in the north-east of the U.S. and comprised of regions 11, 12, 21, and 41[7]. On the other hand, the other camp has a variance of whites being higher than the variance of non-whites for most of the 1980-2000 period. This camp is mainly in the center-south and west of the U.S. and comprised of regions 22, 31, 32, and 33[8].

Note that that extremely high variance, skewness, and kurtosis, for whites in 2000 that can be seen on figures 3.17, 3.19, and 3.21, is from region 22 (i.e. West North Central Division). This aligns with our finding in the previous chapter (by state analysis) where those three measures of risk in 2000 were particularly high for both Minnesota and Kansas (both included in region 22). It adds to our previous analysis by showing this odd increase in "risk"[9] in 2000 for those two states was borne by whites while non-whites seemed unaffected.

On a national level, skewness is overall pretty similar for whites and non-whites (figure 3.18). The large gaps between whites and non-whites' skewness in 1980 and in 2000 is completely driven by outliers. Specifically, the gap in 2000 is driven by the unusually high skewness in the distribution of white individuals in the West North Central Division (region 22). And the gap in 1980 is driven by two different regions: the Middle Atlantic Division (region 12) which had unusually high skewness for non-whites, and the Pacific Division (region 42) which had unusually low skewness for whites. All three outliers can easily be seen in figure 3.19 which show skewness by race (solid black lines for non-whites and dashed red lines for whites) and by region over time.

The smaller (national) gaps from 1990, 2005, and 2015 (figure 3.18), however, stand from the combination of several regions where the skewness of non-whites was

---

[7] Note that region 21 and 42 have one exception each where whites' variance is higher than non-whites' variance: year 1995 and year 1980 respectively.

[8] Note a few exception to this. The variance of non-whites is higher than the variance of whites in 1985 for all of them, in 1980 for regions 31, 32, and 33, and in 2000 for region 33.

[9] Note, however, that theoretically a higher skewness is not equivalent to a higher risk. Our empirical results, though, show positive compensation for higher skewness.

Figure 3.18: Skewness by Race (nationally)

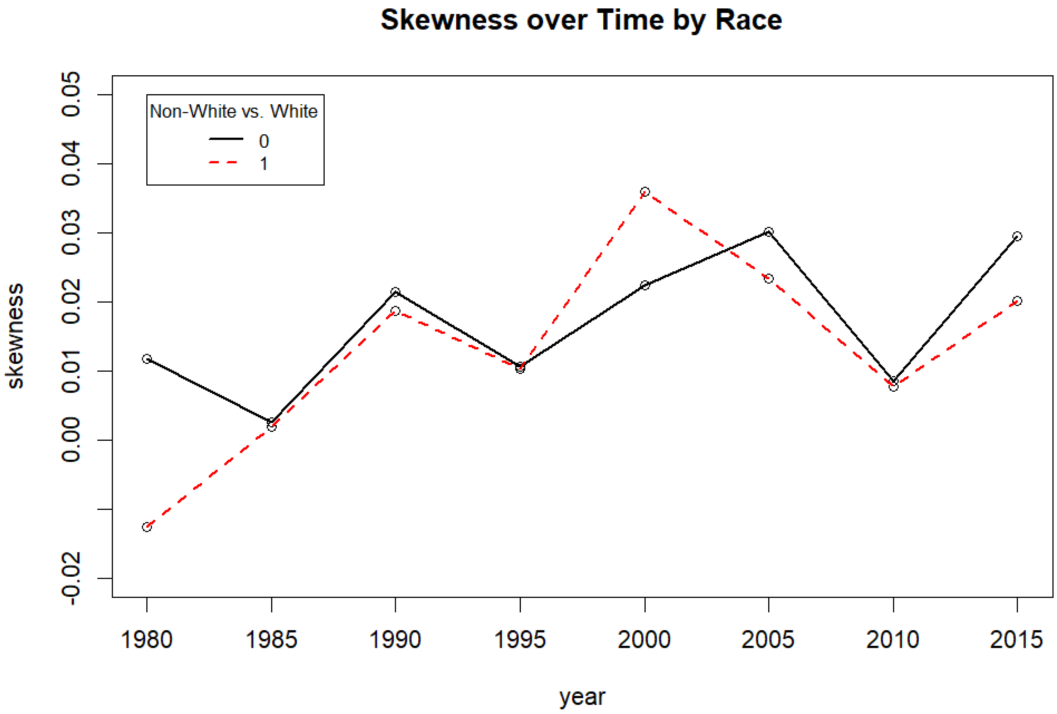


Figure 3.19: Skewness by Race (by region)

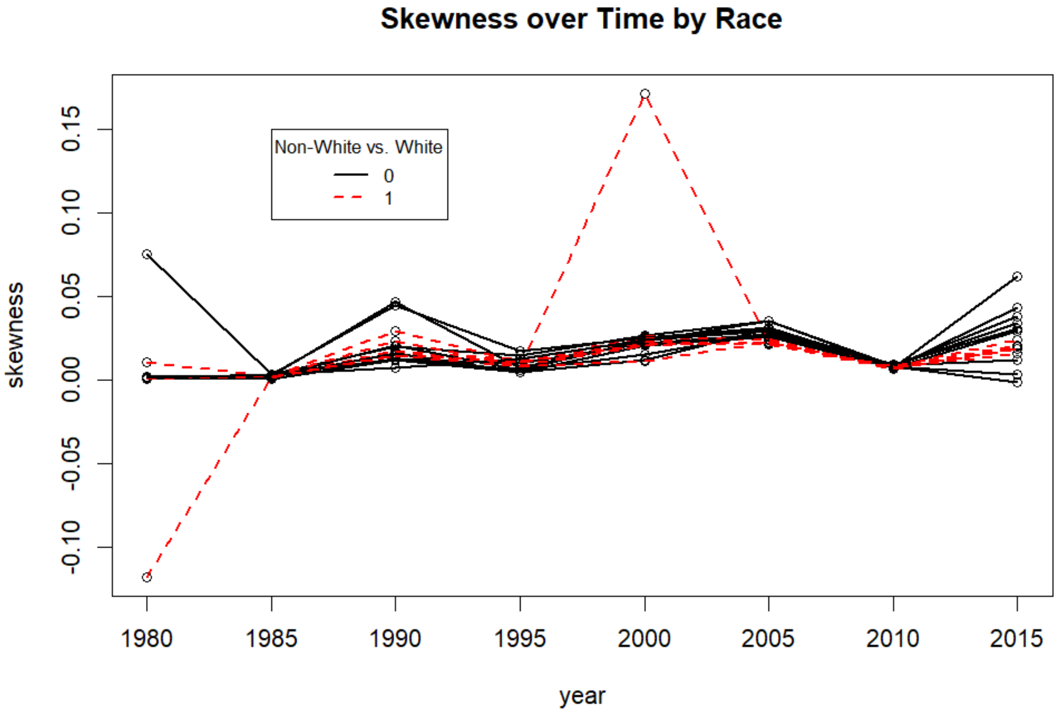


Figure 3.20: Kurtosis by Race (nationally)

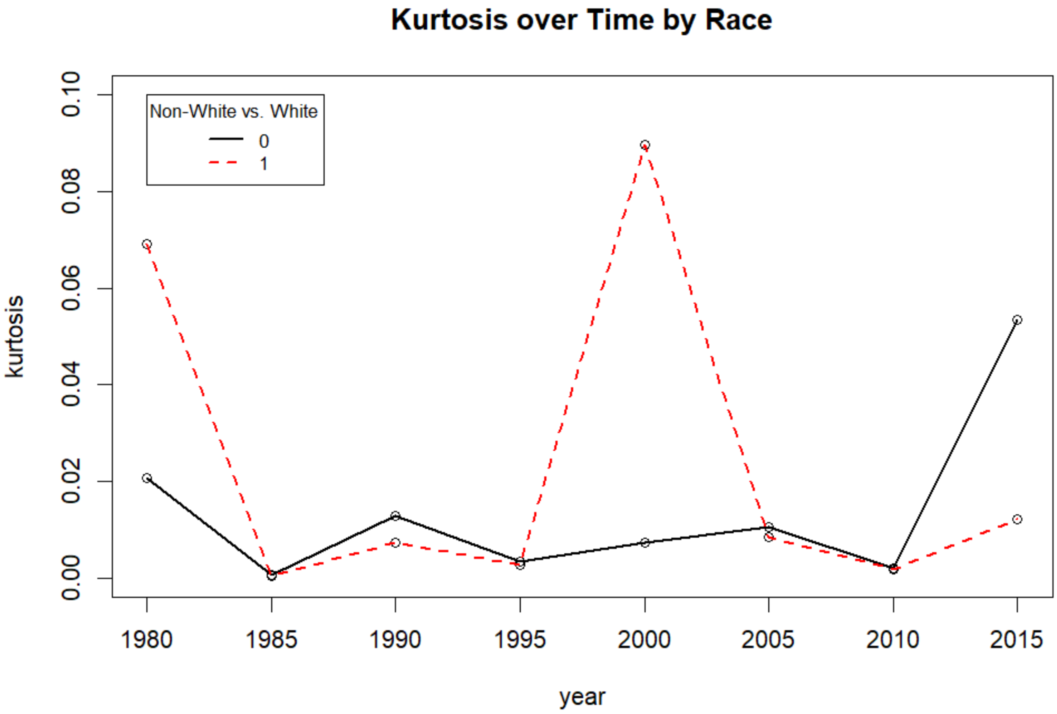
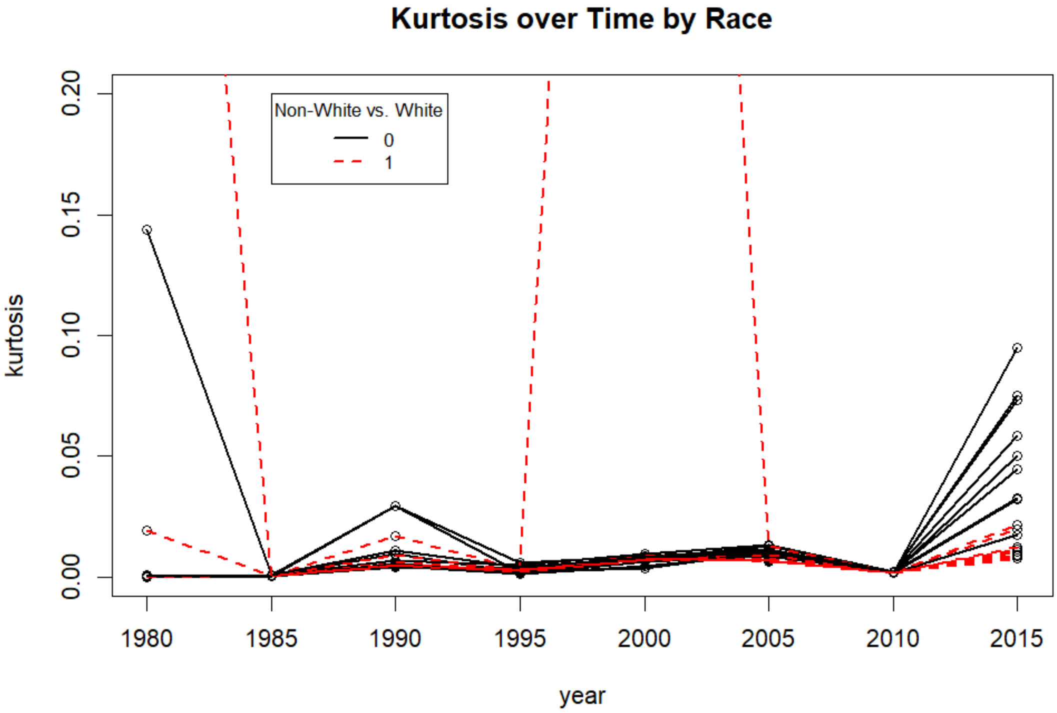


Figure 3.21: Kurtosis by Race (by region)



higher than the skewness of whites. Specifically, in 1990 the small gap is driven by relatively large gaps in regions 12 and 21 (Middle Atlantic Division and East North Central Division), while the other regions did no show a difference in skewness between whites and non-whites. In 2005, the effect is driven by a consistently slightly higher skewness for non-whites than for whites for all regions. Lastly, the regional effects seen in 2015 are mixed. Particularly, the north-east regions 11, 12, and 21, had lower non-whites' skewness than whites' skewness while all the other regions had higher non-whites' skewness than whites' skewness. The biggest gaps of the latter effect being in the Mountain Division (region 41) and then all three southern regions (regions 31, 32, and 33). Note that the regional variation in skewness in 2015 is completely driven by non-whites while whites' regional distributions have very similar level of skewness.

As similar story applies to the national-level kurtosis by race seen in figure 3.20, where the large gaps and values in 1980 and 2000 are driven by a few extreme outliers, the gap in 1990 is driven by a few regions, and the gap in 2015 is driven by most regions. In 1980, the very large kurtosis for the whites' distribution is driven by an extremely large kurtosis for whites in region 42 (Pacific Division) and a slightly high one in region 22 (West North Central Division), while the large but much smaller national-level kurtosis for non-whites is driven by an unusually high kurtosis for non-whites in region 12 (Middle Atlantic Division). To put it into perspective, this means that individuals in those three groups in 1980 had both a high chance of having rate of return to education close to the mean and a high chance of having a rate very far away from the mean (in the tails) but a low chance of being between the mean and the tails (compared to others). Again, according to theory, temperate individuals will prefer to have a lower kurtosis i.e. to avoid the chance of being in the tails. As described previously, the very high kurtosis for whites of 2000 is driven by region 22, while all other regions had similar whites' and non-whites' kurtosis.

While most regions in 1990 did not have a racial disparity for the kurtosis measure, three regions had a kurtosis that was higher for non-whites than for whites. Specifically, the Middle Atlantic Division, the East North Central Division, and the Pacific

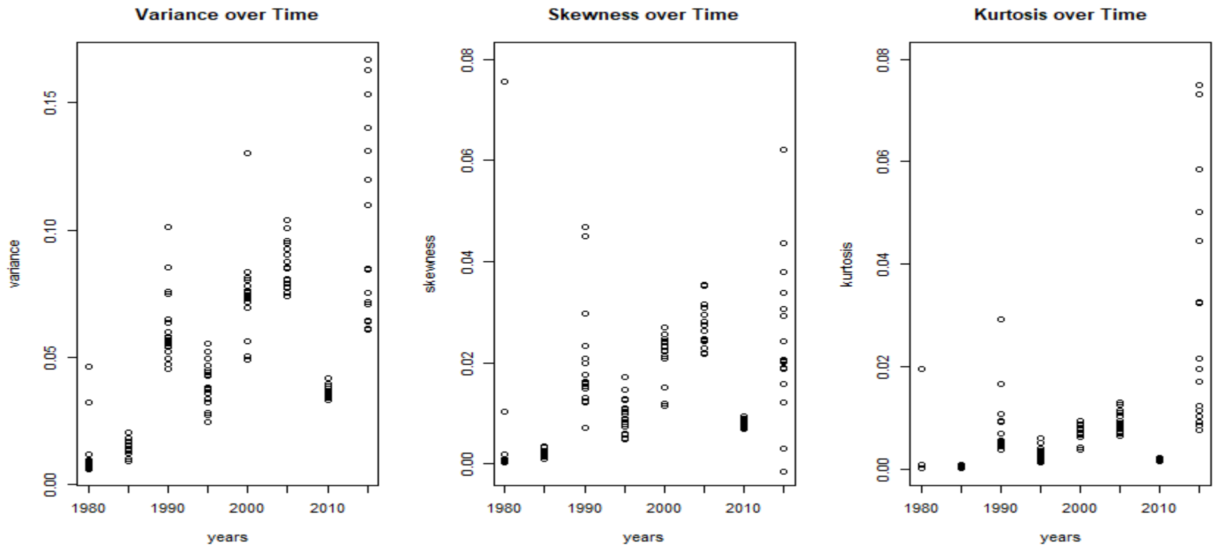
Division (regions 12, 21, and 42). That is, in those three regions, non-white individuals had a higher chance of being in the tails than white individuals. Combined to the similar disparity in skewness in 1990 for the Middle Atlantic Division and the East North Central Division, this means that those non-white individuals had a much higher chance having a really high returns to education (compared to their peers) than white individuals. While this could be considered a "good" type of risk to have to face, it also reveals high inequalities within those groups.

The big national-level gap of kurtosis in 2015 between whites and non-whites is driven by a national effect. That is, all regions apart from one (region 22) have a kurtosis that is higher for their non-whites' distribution than for whites' distribution. While the kurtosis of whites between regions is similar, there is a lot of variation in the kurtosis of non-whites between regions. The two highest kurtosis in 2015 being in the north-east (region 12 and 11 respectively). While this might look like a "good" risk when combine with positive skewness, it is an extremely "bad" risk when combined with a negative skewness (i.e. a high chance of being in the low tail) which is the case for non-whites in region 11.

Overall, the regions in the north-east of the United States, specifically, the New England Division (11), the Middle Atlantic Division (12), and the East North Central Division (21), all stand out with their very high risk - as measured by all three higher moments - faced by non-whites in 2015. The non-whites distributions for all three regions have the top 3 variance and kurtosis and bottom 3 skewness (of all regions and race) in 2015. Hence, the financial crisis seems to have greatly increase the "risk" faced by non-whites in those regions but also increased the inequality both within non-whites and between whites and non-whites.

Zooming in on all three higher moments over time (i.e. excluding outliers) in figure 3.22, we notice they all generally increase over time in a similar way. That is, all three moments start low in the 1980s, subsequently increase greatly and spread out between groups in 1990, momentarily decrease in 1995 to then increase again up until 2005. All have been affected by the 2008 crisis and greatly decrease in 2010 both in

Figure 3.22: Moments over Time by Race and Region (zoom in)



value and in variation between groups. Finally, all three moments seem to completely explode in variation between groups in 2015. Again, this indicates that the 2008 financial crisis momentarily decreased inequality of returns to education to then greatly increase it on the long run both within and between our region-race groups specification.

Using the four moments of our 144 region-race distributions, we nonparametrically (CRS) regress the expected return (mean) on its higher moments controlling for race (white dummy) and region fixed effects. The resulting gradients on the three higher moments are summarized in table 3.9 for all gradients and in table 3.10 for statistically significant gradients only. Table 3.11 presents the equivalent OLS results.

The main difference between the linear regression and the nonparametric regression here is the effect of skewness on expected returns. While the OLS results suggest that a higher skewness will be compensated by a higher rate of return, the CRS results suggest the opposite (the few statistically significant gradients are all negative). Note that the CRS results also indicate that skewness has very little effect on expected returns as only 16 of the 144 gradients are statistically significant. On the other hand, more than 60% of gradients on variance are statistically significant and all of the gradients on kurtosis are statistically significant. Overall, those results seem to indicate that the relationship between expected rate of returns to education and its higher orders greatly depends on how we divide individuals.

Table 3.9: Spline Gradients Summary Statistics (by race and region)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
g.variance	-1.67	0.39	2.19	2.06	3.26	7.70
g.skewness	-0.22	-0.09	-0.04	-0.02	0.07	0.17
g.kurtosis	-0.12	-0.11	-0.09	-0.09	-0.08	-0.07

Table 3.10: Spline Gradients Summary Statistics (by race and region): Statistically Significant Results Only

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	obs.
g.variance	-1.48	2.20	2.91	2.93	3.97	7.70	89
g.skewness	-0.22	-0.22	-0.21	-0.21	-0.21	-0.21	16
g.kurtosis	-0.12	-0.11	-0.09	-0.09	-0.08	-0.07	144

Table 3.11: OLS Results (by race and region)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.09	0.01	8.34	0.00
variance	1.42	0.13	10.73	0.00
skewness	0.48	0.17	2.93	0.00
kurtosis	-0.20	0.04	-5.34	0.00
white	0.01	0.01	2.44	0.02
time	0.00	0.00	2.49	0.01

Figure 3.23 compares the gradients' density between white (red dashed lines) and non-white individuals (thin black solid line). The thick black solid line on the left panels and the histograms on the right panels represent the gradients' distributions of both groups (whites and non-whites) combined. Interestingly, compensation for variance tends to be higher for white individuals than for non-white individuals while the opposite is true for compensation for skewness. Specifically, non-white individuals seem to be less (or not at all) compensated for variance as more non-white distributions have a gradient on variance close to zero. For kurtosis, white individuals tend to be less hurt - through a lower expected return - by higher kurtosis compared to non-white individuals.

Figures 3.24 - 3.26 show the gradients on each moments against expected rate of returns to education (left panels) and against their respective moments (right panels). On each graph, the black circles represent non-whites' distributions and the red

Figure 3.23: CRS Gradients Density and Histogram by Race (and by region)

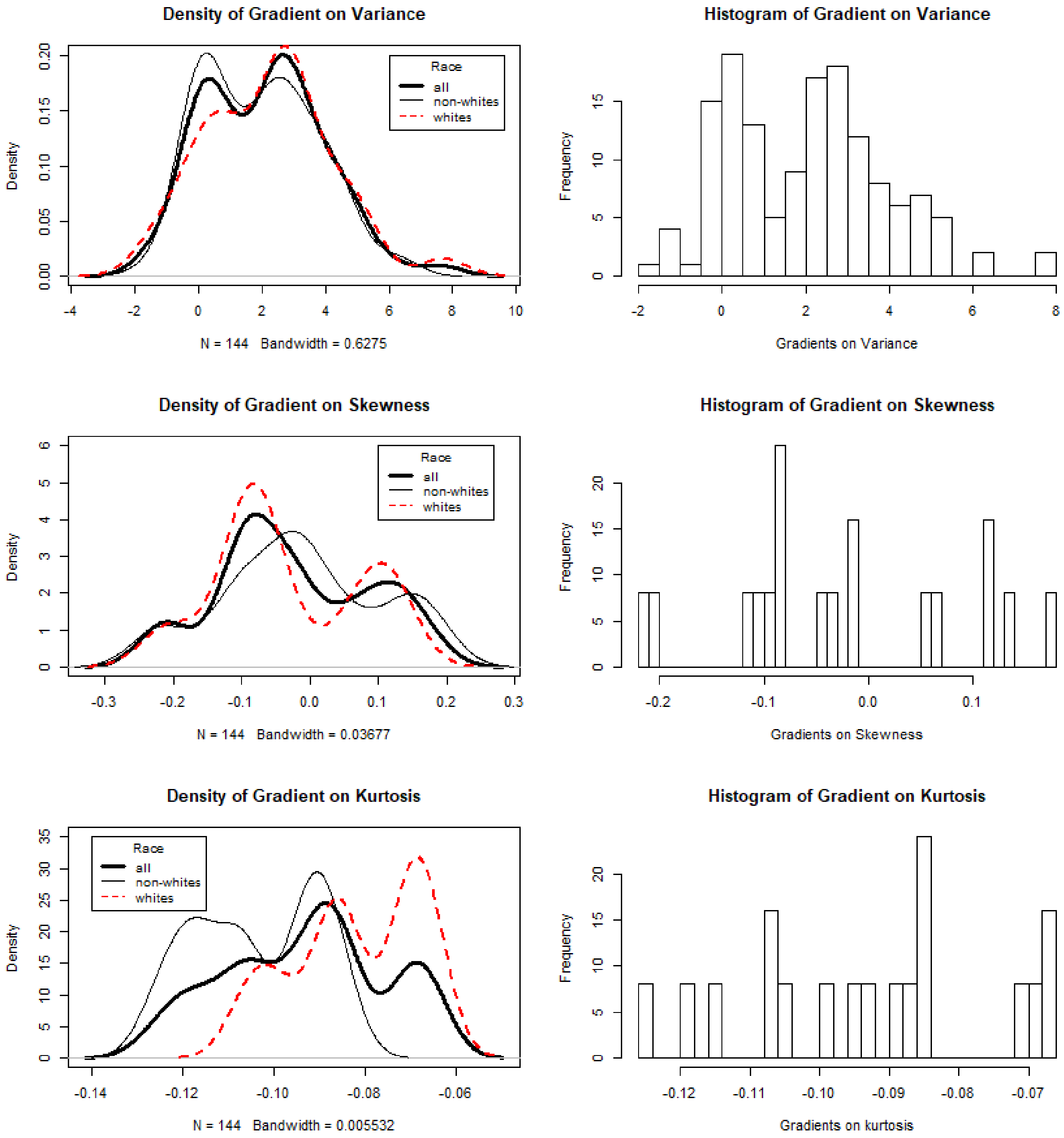


Figure 3.24: Spline Gradients on Variance (Stat. Sign. Only)

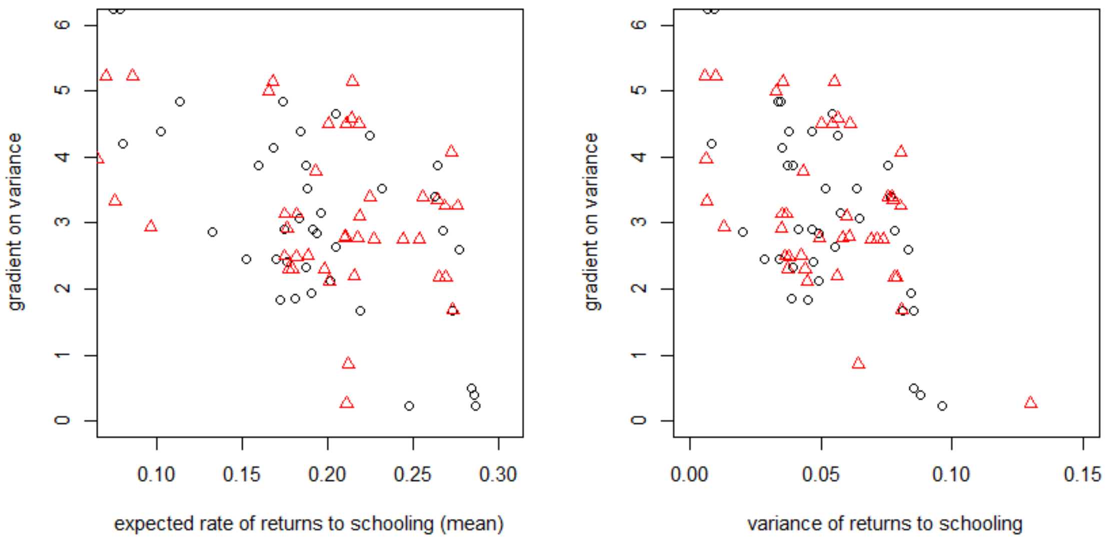


Figure 3.25: Spline Gradients on Skewness (Stat. Sign. Only)

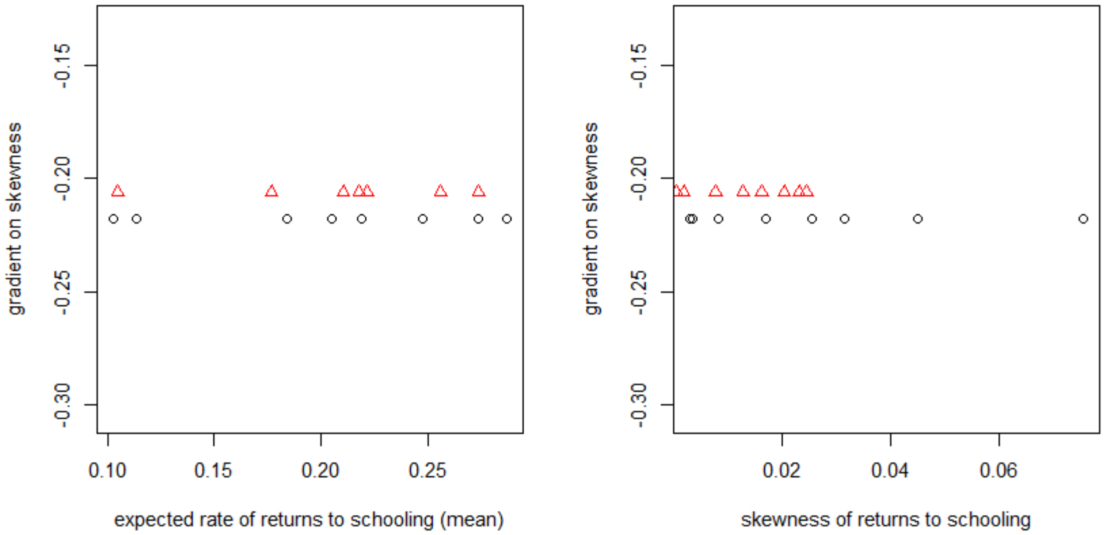
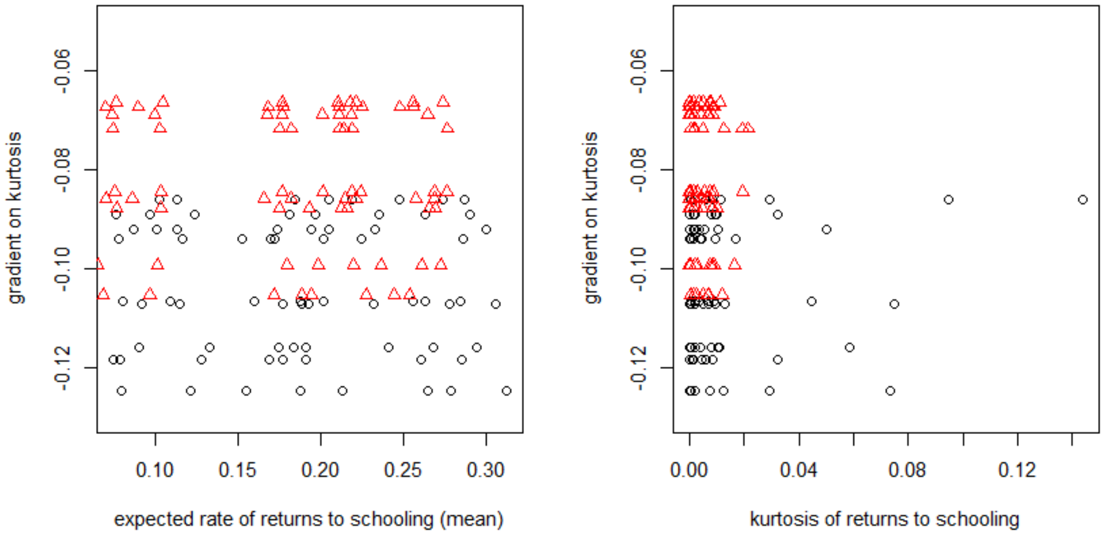


Figure 3.26: Spline Gradients on Kurtosis (Stat. Sign. Only)



triangles represent whites' distributions. Also note that only the statistically significant gradients are shown.

While the full 144 gradients on variance do not seem to show any particular relationship when plotted against expected returns and variance, focusing on the 89 statistically significant gradients only show a negative relationship between each pair. That is, distributions with higher expected returns tend to show less compensation for variance. Similarly, distributions with low variance are at the margin compensated much more generously for an increase in variance than distributions with higher variance to begin with. As is also shown by our other analysis, it seems that after a certain point, i.e. high enough variance, increasing variance further is not compensated by higher expected returns[10].

As shown in figure 3.25, only a few gradients on skewness are statistically significant and they are all negative. There seems to be no connection between the magnitude of skewness and the likelihood for its gradient to be statistically significant. Interestingly, however, all of the statistically significant gradient on skewness are in region

---

[10] Note that while we control for time with a time trend which has a very small positive coefficient, the compensation for variance still seems to decrease over time. However, the compensation for skewness and kurtosis do not vary in a particular way over time.

12, i.e. the Middle Atlantic Division. Meaning, skewness only seems to be (negatively) compensated for in that region. Though it seems that all of the non-white distributions have slightly higher (in magnitude) gradients on skewness than white distributions, the difference is very mild. Specifically, an increase in skewness of 0.01 units will result in an decrease in expected return to education of .22 percentage points for non-whites and .21 for whites (in the Middle Atlantic Division).

The gradients on kurtosis shown in figure 3.26 are all statistically significant and negative. There does not seem to be any particular relationship between (negative) compensation for kurtosis and expected rate of return to education or between compensation for kurtosis and kurtosis itself. However, distributions of white individuals (red triangles) are consistently negatively compensated by kurtosis to a lesser extend than distributions of non-white individuals. On average, when the kurtosis of white men's distributions increase by 0.01 units, their expected rate of return to education will decrease by 0.08 percentage points (0.0008 units). Whereas, when the kurtosis of non-white men's distributions increase by 0.01 units, their expected rate of return to education will decrease by 0.10 percentage points (0.001 units) on average.

3.3.3 Analysis by Education-Level

In our final analysis we subset individuals' rate of return by region and by education-level. We create three different levels of education: no high-school degree (no HS), high-school degree (HS), and college degree. The 9 regions used are the same as the analysis by race and described in table 3.8.

Figure 3.27 shows the expected rates of return over time by education level. Unsurprisingly, individuals without a high-school degree (black solid lines) can expect lower rates of return to one more year of education than individuals who hold a high-school degree (red dashed lines) throughout the whole period. Similarly, holding a college degree (green dotted lines) tends to result in higher expected returns than holding a high-school degree only. Because those rates of return represent the return to one more year of education on average, this means that the returns from one year of college are generally much higher than the returns from one year of high-school. While

Figure 3.27: Expected Returns by Education-Level

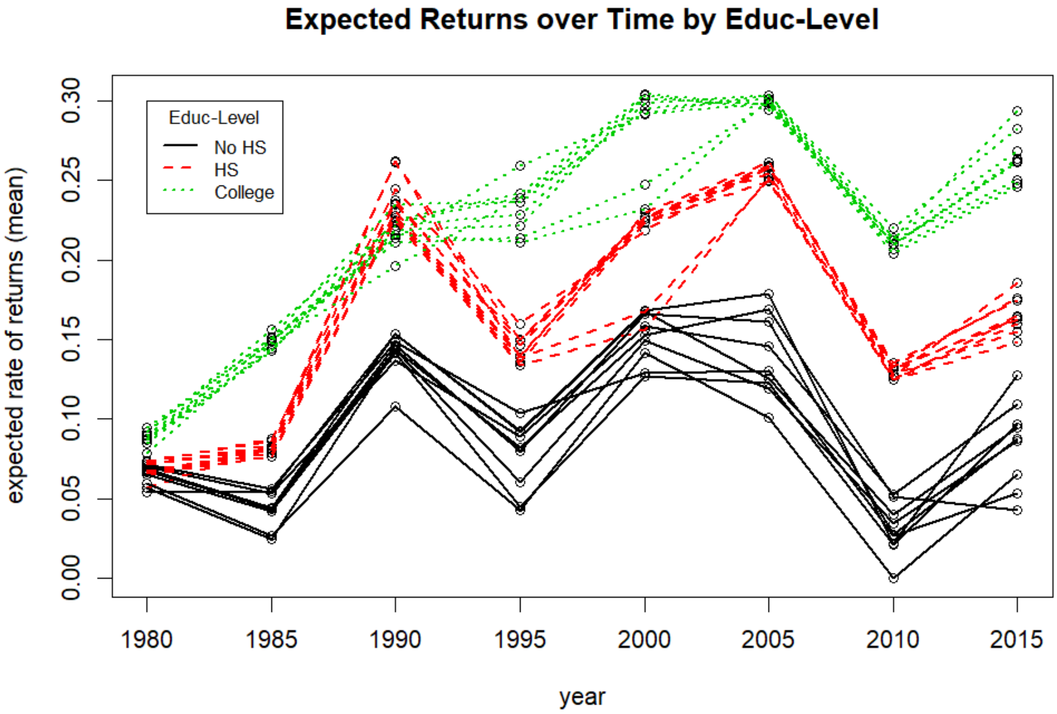
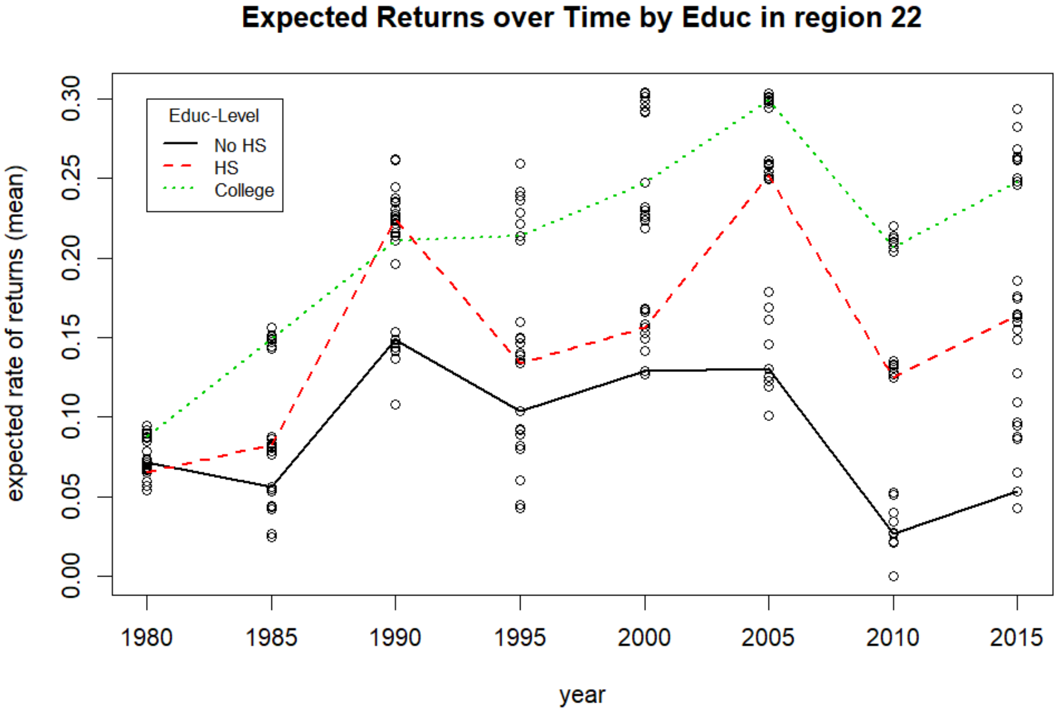


Figure 3.28: Expected Returns by Education-Level (West North Central Division)



the gaps between each education-level were very small in 1980, they grew quite fast and stayed large until the end of the period.

The only exception to this education gap being in 1990 where high-school graduates have on average higher rates of return to education than college graduates. As you can see on figure 3.27, the expected returns of college graduates had a continuous increasing trend between 1985 and 1995 while the expected returns of individuals without a college degree sharply increased between 1985 and 1990 and then sharply decreased between 1990 and 1995. This effect matches our analysis by occupation seen in figure 3.2. That is, occupations with generally lower expected returns and lower levels of education required like farming (occupation 19 on lower panel of figure 3.2) see a sharp increase in expected returns between 1985 and 1990 followed by a decrease. On the other hand, the expected returns of occupations generally requiring more education (higher degree), like legal jobs (occupation 8), either slightly increased or stagnated both between 1985 and 1990 and between 1990 and 1995.

While there is a lot of variation in expected returns between regions for men without a high-school diploma (i.e. the black lines are relatively spread out), expected returns for high-school and college graduates are more similar between regions (i.e. dotted green lines and dashed red line tend to be respectively tight together). Two regions, however, break away from the pack in 2000 with much lower expected rates of return: the West North Central Division (region 22 shown separately in figure 3.28) and the Mountain Division (region 41). While the expected returns for high-school and college graduates in those two regions increased more slowly than other regions between 1995 and 2000, they subsequently catch up to the other regions by 2005. As a result, for both regions in 2000, college graduates have expected returns that are closer to the expected returns of high-school graduates from other regions. Similarly, high-school graduates in those two regions have expected returns that are closer to the expected returns of men without a high-school diploma in other regions. Generally, within each education-level, this variation between regions seems to have increased slightly in the 1990s and in 2015 compared to other years.

Figure 3.29: Variance by Education-Level

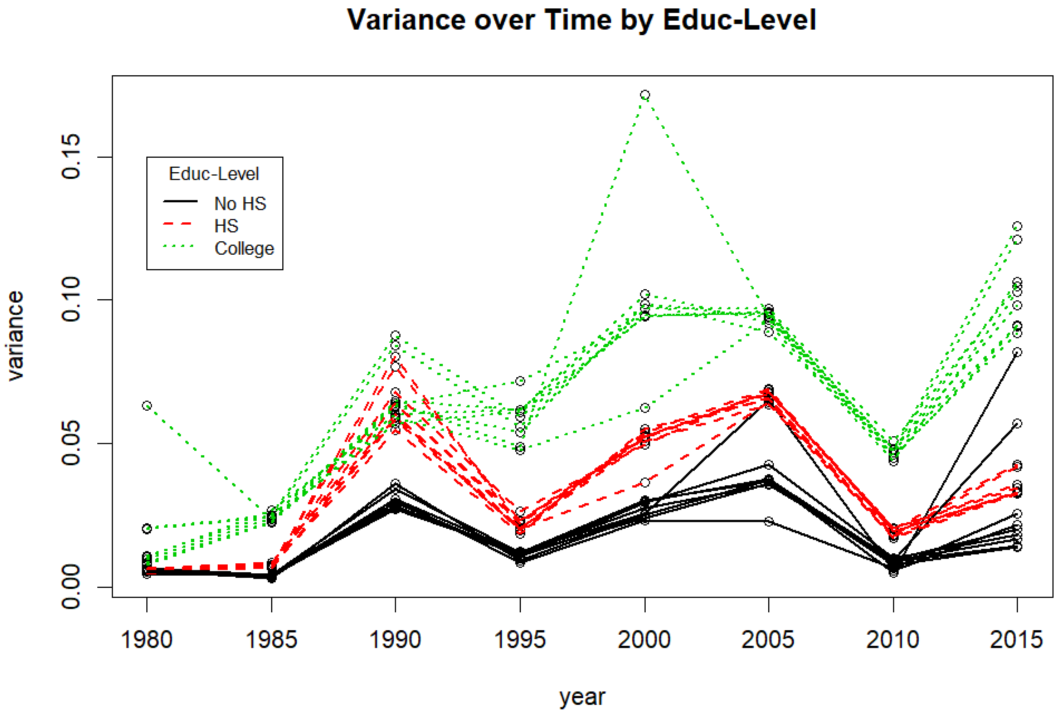


Figure 3.30: Skewness by Education-Level

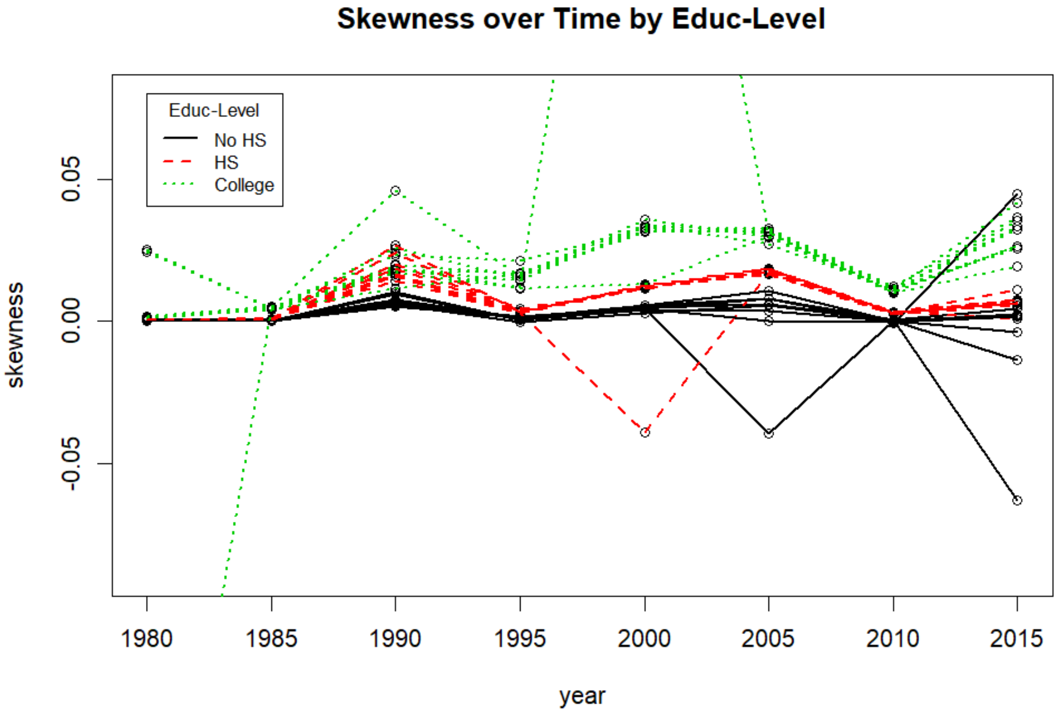


Figure 3.31: Kurtosis by Education-Level

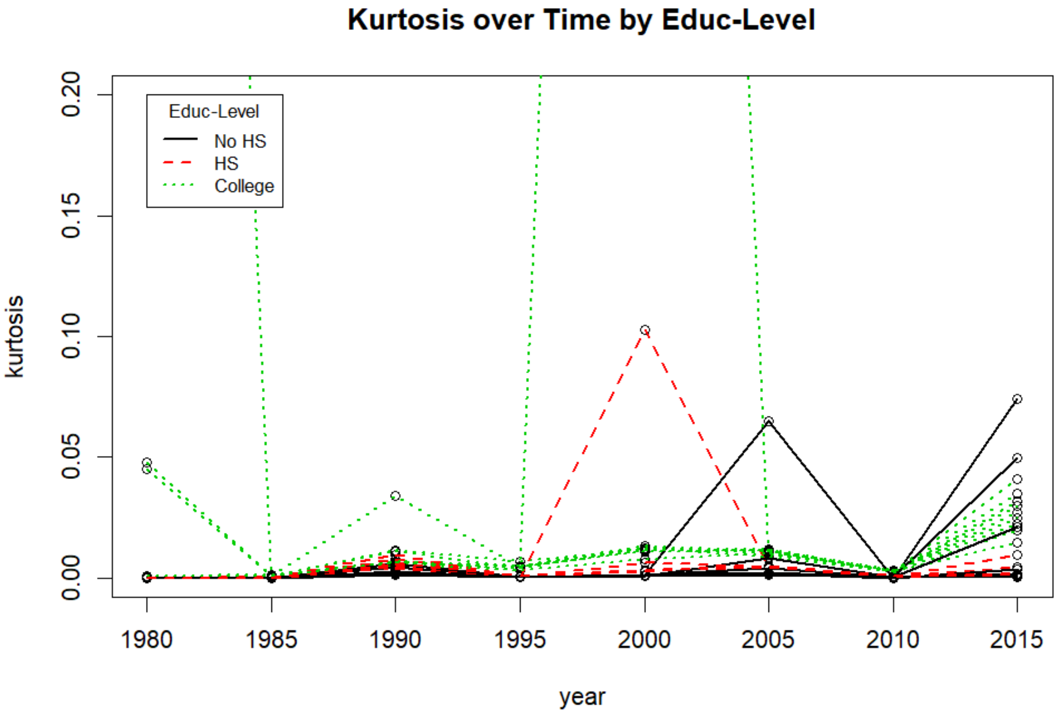
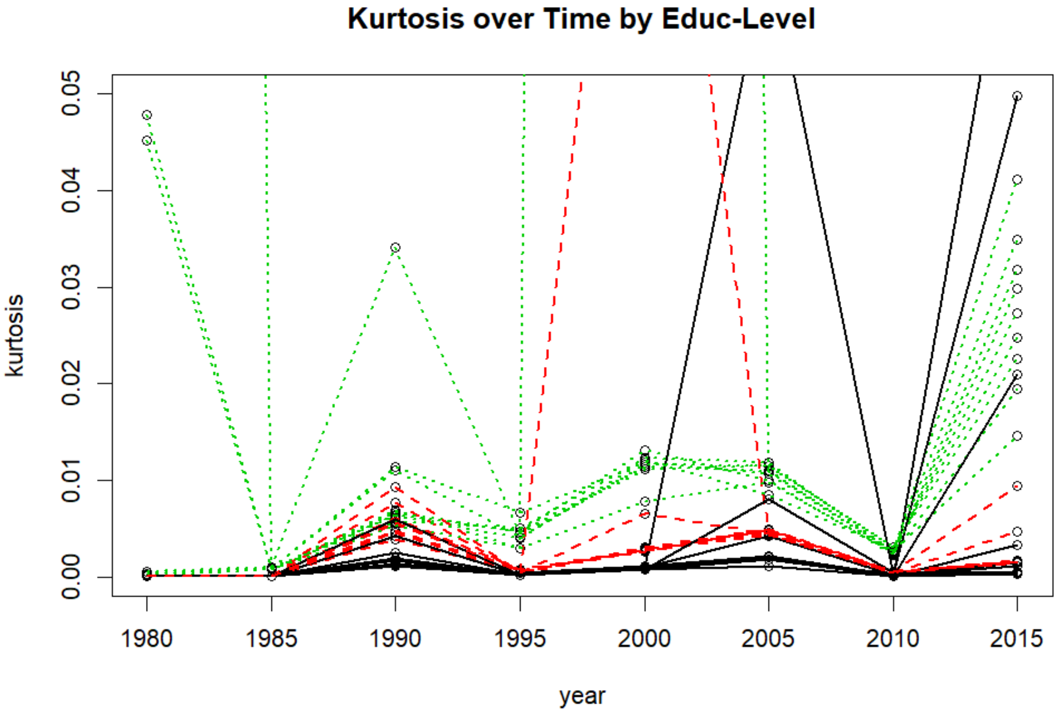


Figure 3.32: Kurtosis by Education-Level (zooming-in)



Overall, the three higher moments shown in figures 3.29 - 3.31 have higher values for college graduates distribution than for high-school graduates distributions and higher values for high-school graduates distributions than for "no high-school" distributions. This indicates that the higher the degree earned, the higher the inequality of rates of return to education within that group[11].

Interestingly, the patterns of the higher moments are similar to those of the expected returns described above. Notably, in 1990, the higher moments become really similar between men with a high-school degree and men with a college degree indicating that the overall shapes of the distributions between the two groups were very similar. Also, the variation in risk (as measured by the higher moments) between regions and within each education-level are higher in 1990 and in 2015 compared to other years. The geographical variation of risk within the "no high-school" group was particularly high in 2015.

Once again, the West North Central Division (region 22) stands out in 2000 for all three higher moments. Specifically, for college educated men, the variance, the skewness, and the kurtosis, all peak to extremely high values. That is, college educated men in the West North Central Division faced an extremely unequal rates of return distribution with a lot of them far away from the mean (high variance) and particularly in the tails (high kurtosis) and most of them were on the low end of the distribution with a few and the extremely high end of the distribution (highly positively skewed). For high-school graduates, the distribution is highly negatively skewed with a very high kurtosis. The variance is, however, similar to other regions in 2000. Finally, men without a high-school degree did not face a particularly different distribution compared to their peers in other regions. They did face, however, higher than usual risk in 2005 with a variance similar to those of high-school graduates, a highly negative skewness, and a high kurtosis (highest kurtosis in 2005). Going back to figure 3.28, note that all three groups experienced very low expected returns compared to their peers in other regions in 2000 suggesting that they were not well compensated for the existing variation

---

[11] Note that we did not standardized the higher moments, hence, it is not surprising for the higher moments to follow that order

within their respective groups.

In 2015, the New England Division (region 11) and the East South Central Division (region 32) also stand out with the particularly high risk faced by men without a high-school diploma (solid black lines). Specifically, their distributions had higher variance than high-school graduates (two higher ”no HS” points in 2015 on figure 3.29) and higher kurtosis than both high-school graduates and college graduates (two highest points in 2015 on figure 3.31). Additionally, their skewness were also extreme with the New England Division having the lowest skewness in 2015 (lowest negative point on figure 3.30) and the East South Central Division having the highest skewness in 2015 (highest point on figure 3.30). Interestingly, the former region had the lowest expected returns for that year and the latter region had the highest expected returns of all the ”no HS” distributions. This suggest that uneducated men in those two regions are experiencing particularly strong variability (risk) at the end of the period which probably stems from long-run effect from the 2008 financial crisis.

Lastly, college graduates in 1980 living in the Pacific Division (region 42) also had particularly extreme higher moments. While their variance was the highest of 1980, their skewness had the lowest value over the whole sample and their kurtosis the highest (on par with region 22 in 2000). In other words, college graduates' rates of return to education in the Pacific Division were spread out with many in the tails, mainly in the higher-end tail. While this shows that they were a lot of inequality within this group, it also indicates that college graduates there faced a particularly high risk in the sense that their rate of return to one more year of education was particularly uncertain.

Table 3.12: Spline Gradients Summary Statistics (by educ-level and region)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
g.variance	0.77	2.01	3.72	5.70	10.19	14.65
g.skewness	-25.13	-1.04	1.11	2.36	4.50	34.15
g.kurtosis	-316.91	-107.25	-24.31	-77.85	-2.59	2.24

To estimate the differences in risk-return to education trade offs between education-level, we use the four moments of our 216 region-educ-level distributions described thus far. Specifically, we regress the mean on its higher moments while controlling for

Table 3.13: OLS Results (by educ-level and region)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.05	0.01	6.97	0.00
variance	2.24	0.10	22.83	0.00
skewness	-0.01	0.08	-0.10	0.92
kurtosis	-0.16	0.02	-8.92	0.00

education-level and region fixed effects. The nonparametric gradients (CRS) are summarized in table 3.12 while the linear coefficients (OLS) are shown in table 3.13. Notably, the nonparametric gradients have generally higher values than the linear coefficients suggesting that OLS underestimates the "compensation" for risk (as measured by the higher moments).

Table 3.14: Spline Gradients Summary Statistics (by educ-level and region):
Statistically Significant Results Only

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	obs.
g.variance	1.18	2.01	3.73	5.73	10.27	14.65	215
g.skewness	-25.13	-2.68	2.95	3.07	8.60	34.15	153
g.kurtosis	-316.91	-128.73	-24.39	-78.95	-2.60	2.24	213

Table 3.15: Spline Gradients Summary Statistics by Education-Level (by region):
Statistically Significant Results Only

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	obs.
g.variance (No HS)	2.04	3.66	10.27	8.09	11.86	13.45	72
g.variance (HS)	1.80	2.01	4.95	6.13	9.48	14.65	72
g.variance (College)	1.18	1.47	2.01	2.92	2.15	8.79	71
g.skewness (No HS)	-2.74	-2.70	4.80	4.06	9.81	18.45	52
g.skewness (HS)	-19.78	-12.60	3.71	1.25	8.17	25.84	52
g.skewness (College)	-25.13	0.97	1.27	3.94	3.42	34.15	49
g.kurtosis (No HS)	-313.72	-252.62	-78.83	-132.75	-34.40	-1.53	69
g.kurtosis (HS)	-316.91	-110.17	-30.54	-87.06	-2.61	2.24	72
g.kurtosis (College)	-241.56	-11.45	-2.64	-19.28	-2.33	-0.36	72

Tables 3.14 and 3.15 show the summary statistics of statistically significant gradients (CRS) only, the latter table breaks down the former by education-level to better show the compensations' heterogeneity. While most of the gradients on variance and kurtosis are found the be statistically significant, the effect of skewness on expected return is less comprehensive with only 70% of its gradients being statistically significant.

Figure 3.33: CRS Gradients Density and Histogram by Educ-Level (and by region)

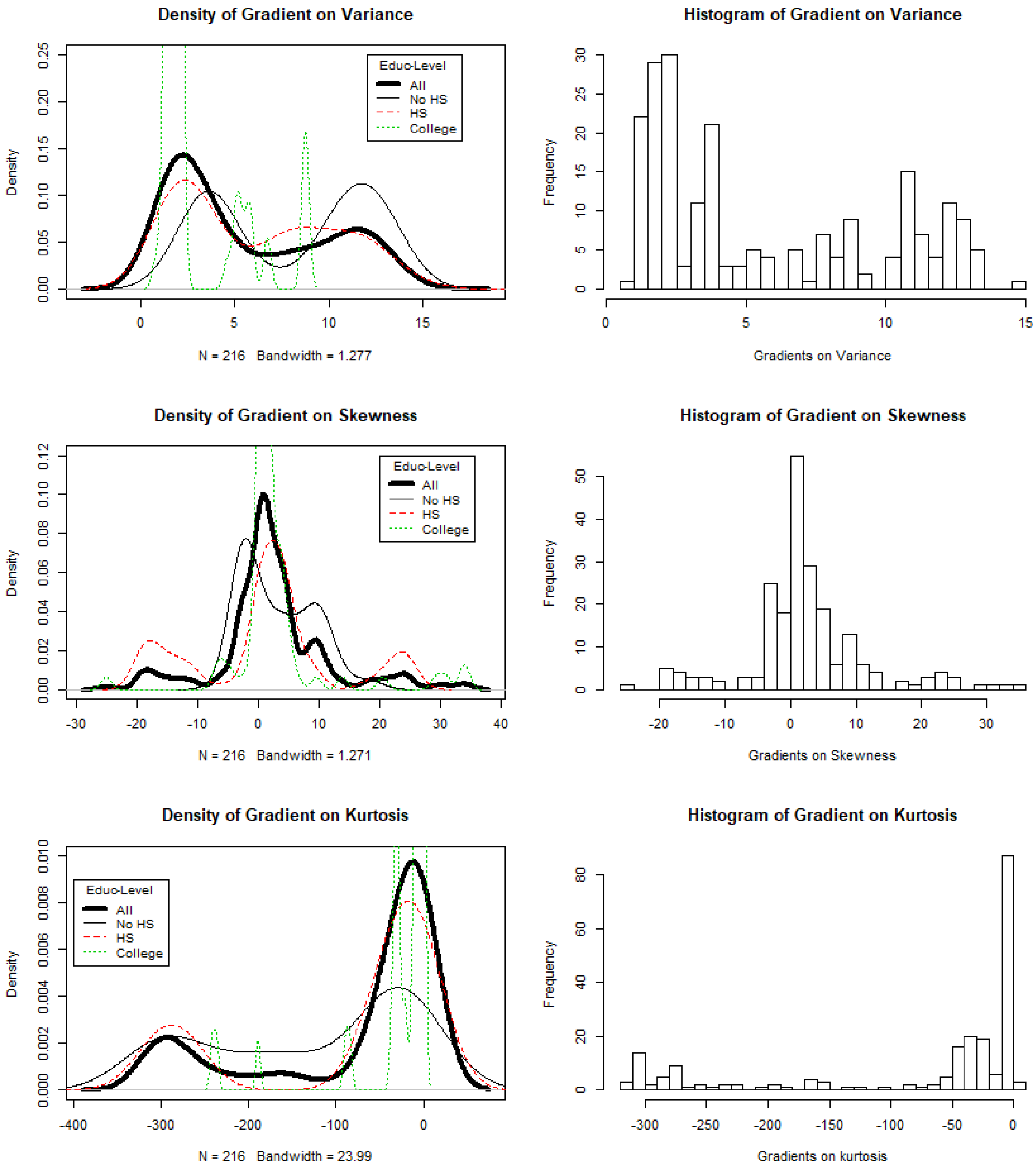


Figure 3.34: Spline Gradients on Variance (Stat. Sign. Only)

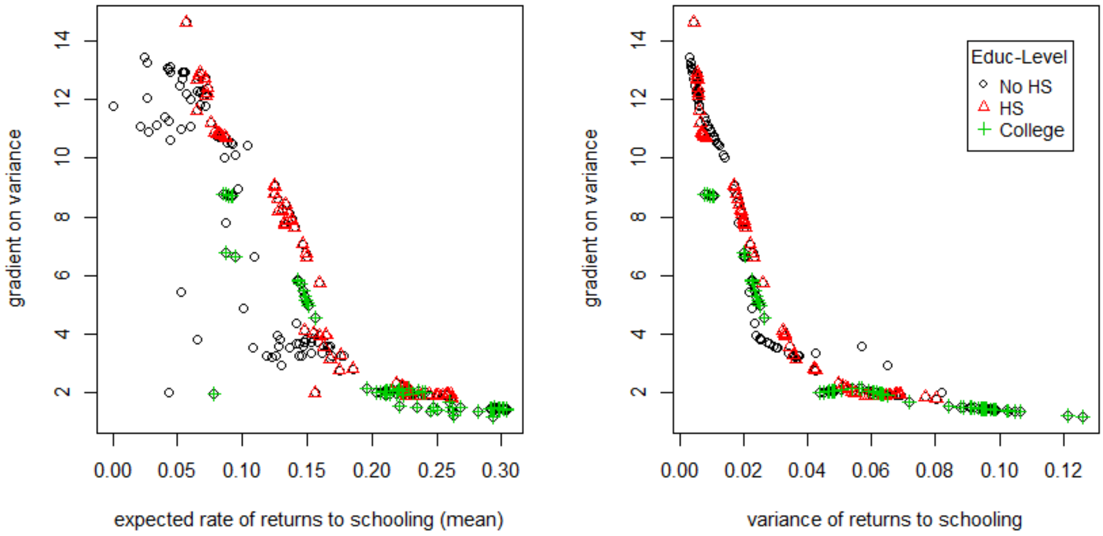


Figure 3.35: Spline Gradients on Skewness (Stat. Sign. Only)

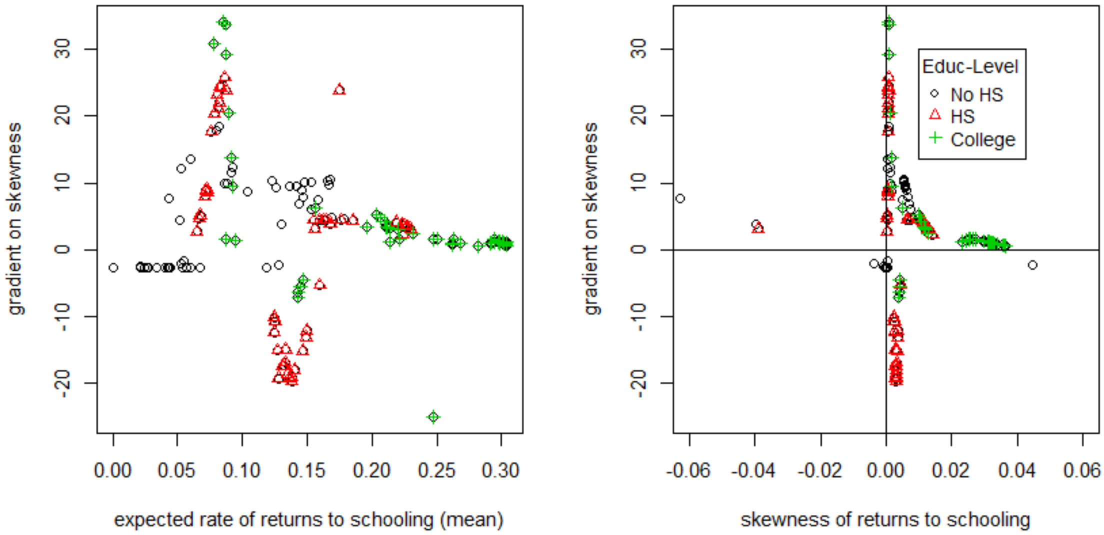
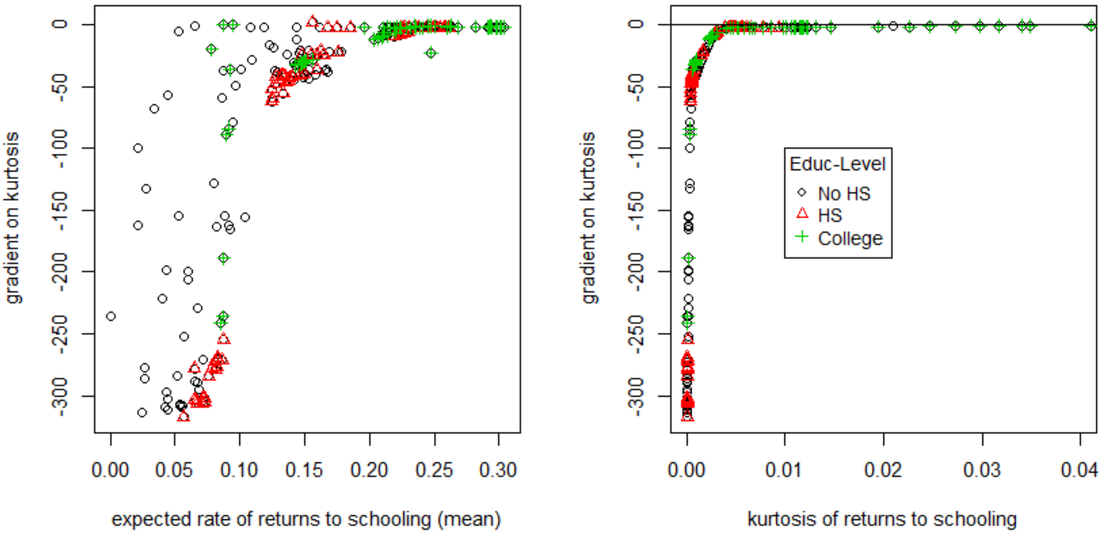


Figure 3.36: Spline Gradients on Kurtosis (Stat. Sign. Only)



There is clear evidence of strong compensation for variance within region-education-level groups. The gradients on variance are all positive and all but one are statistically significant. The first three rows of table 3.15 show that variance in rates of return within uneducated (no high-school degree) men is better compensated than variance within high-school graduates and variance within high-school graduates is overall better compensated than variance within college graduates. In particular, for an increase in variance of 0.01 units, uneducated men see on average an increase of 8.09 percentage points in their expected return (i.e. 0.0809 units) while high-school graduates' expected return increase by on average 6.13 percentage points and college graduates' expected return increase by on average 2.93 percentage points. Figure 3.34 shows that both the relationships between the gradients on variance and expected returns and the gradients on variance and the variances are negative. That is, everything else equal, as the variance of rates of return to education within region-education-level groups increases, the compensation for that variance decreases. Because uneducated men tend to have lower variances compared to educated men (HS and college), it explains why they are better compensated for variance.

The compensation for skewness is less straightforward. The relationships between the gradients on skewness and both expected returns and skewness itself have heartbeat-like shapes (see figure 3.35). In particular, for extremely low positive skewness (between 0 and 0.0015) the gradients are positive and large, while between skewness values of about 0.0015 and 0.005 the gradients become negative with relatively large values. However, starting from a skewness of 0.005, the gradients become positive again with values around 10 and decrease in magnitude as skewness increases. Note that there is proportionally a lot more college graduates distributions in this latter group and a lot more high-school graduates distributions in the second group (with negative gradients). The few distributions with large negative skewness have positive gradients on skewness while those close to zero have negative gradients. Overall, the compensations' magnitudes are relatively high[12] and decrease increasingly as the skewness gets further away from zero. This effect can also be observe with the magnitudes of the gradients on variance and kurtosis.

The great majority of the gradients on kurtosis are statistically significant and negative. This suggests, contrary to theory on decision making under uncertainty, that individuals making educational investment decisions favor higher kurtosis and are thus negatively compensated via a lower expected returns for higher kurtosis. All of the extreme negative gradients are on extremely small kurtosis. As kurtosis gets higher the gradients converge to zero very fast as shown on figure 3.36. For example, around a kurtosis of 0.01, increasing the kurtosis by 0.001 (a more reasonable value considering the range of kurtosis) will result in a decrease in expected returns of about 0.14 percentage points (i.e. gradient of -1.4). From table 3.15 you may notice that the gradients on kurtosis are generally smaller in magnitude for more highly educated men. This stems from the nature of the heterogeneity of kurtosis between education-level groups. That is, kurtosis tend to be higher for men with higher degrees.

---

[12] For example, increasing the level of skewness by 0.01 units would result in an increase in the expected rate returns of 3.07 percentage points on average.

3.4 Conclusion

This paper extends on the previous chapter by sub-sampling the rates of return to education in three different ways to measure risk: by occupation, by region and race, and by region and education-level. That is, the shape of those distributions respectively measure the particular risk of working in a specific occupation, of working in a specific region and being white vs. non-white, and of working in a specific region and holding a degree vs. not. The main aim of the paper being to investigate the heterogeneity of risk and compensation for risk between those groups.

In the analysis by occupation, we find that low paying jobs tend to have lower risk - as measured by the three higher moments - and less variation in risk from year to year. This means that men working in those occupations tend to all have very similar rates of return to education. However, the expected return to education for men in those occupations varies a lot more from year to year than for men working in higher-income jobs. This suggest that low paying jobs like farming, fishing, forestry, or extraction jobs, are a lot more affected by the health of the general economy. For example, the early 1990s recession seems to have affected those jobs more drastically than other jobs.

In the analysis by race, we find that expected returns for non-whites are slightly higher than for whites between 1980 and 2015 with the exception of the 1990's. The risk as measured by the variance appears very similar on average (i.e. nationally) between whites and non-whites between 1980 and 2000. From 2005 onwards though, non-whites start to have a higher variance than whites. This racial disparity momentarily decreases in the aftermath of the 2008 financial crisis but greatly exacerbates a few years later (2015). This indicates that starting in the 2000's, non-white men in the U.S. started to face greater inequality in terms of returns to education than white men. On a national level the skewness and kurtosis also tend to be higher for non-white men than for white men, the few exceptions being driven by a few regional outliers.

When looking at the racial disparity in returns to education distributions by region, we note a lot more variation in the distributions of non-white men between regions than in the distributions of white men between regions. That is, white men tend to face relatively more similar distributions across the different U.S. regions compared to non-white men. Overall, the regions fall into two groups. In the north-east of the U.S., the returns' distribution of non-white men tends to face higher variance than non-white men for the entire 1980-2015 period. On the other hand, for the center-south and west of the U.S., the variance of the distributions of white men tend to be higher than the variance of non-white men for the 1980-2000 period. Between 2005 and 2015, non-white men systematically face more risk as measured by both the variance and the skewness than white men.

Specifically, in 2015, the regions in the north-east of the U.S. particularly stand out (i.e. the New England Division, the Middle Atlantic Division, and the East North Central Division) with unusually high risk - as measured by all three higher moments - faced by non-whites. The rates of return's distributions of non-white men for all three north-east regions have the top three variance and kurtosis and bottom three skewness of all region-race observations in 2015. In other words, non-white men in those three north-east regions faced exceptionally high risk at the end of the sample period.

The analysis by region and education-level predictably shows that uneducated men have generally lower rates of return to education than educated men. That is, on average, a year in college yields much higher returns than a year in high-school. Similarly, someone who finished high-school has on average a higher rate of return on one year of schooling than someone who didn't finish high-school. Interestingly though, the expected return of uneducated men varies a lot more between regions than the expected return of educated men (HS and college). Hence, where you live influences your return to education a lot more if you do not hold any degree.

Generally, all four moments of the region-education-level distributions have similar patterns over time. For example, the higher the education level the higher the value of the four moments. This indicates that the higher the degree earned, the higher the expected returns but also the higher the inequality of rates of return to education within that education-level group. In 1990, the four moments become really similar between men with a high-school degree and men with a college degree indicating that the overall shapes of the distributions between the two education-level groups were very similar during that year. In other words, it seems that holding a college degree as opposed to only a high-school degree did not make much of a difference in terms of returns in 1990. Also, similarly to the previous analysis, there are particularly high geographical variability of risk in both 1990 and 2015, as well as in 2005 for uneducated men specifically.

In all three analysis, we find that the 2008 financial crisis momentarily reduced the respective "risk" faced by men, i.e. all of the distributions became very similar in terms of shape and magnitude. However, the crisis seems to have increased their risk in the long-run as inequality both within and between the different groups of men analyzed has increased.

Unsurprisingly, years that showed less variation in risk between groups like 2010 had the least amount of statistically significant risk compensation. That is, when everyone in the population face similar risk levels, individuals stop being compensated for that risk.

From regressing the mean on the higher moments, we find for all three analysis strong evidence of compensation for all three types of risk. The evidence is especially strong for the specific risk (i.e. variation) existing within a particular education-level with most of the gradients being statistically significant. Generally, we find that skewness matters less than kurtosis. In other words, individuals are more systematically compensated for a change in the kurtosis of their group's distribution than for a change in the skewness. Also, contrary to theoretical predictions, a higher kurtosis within a particular group results in a lower expected return. This suggest that individuals are more concern about compensating for a lower peak (i.e. lower chance of being close to the mean) than they are about compensating for fatter tails (i.e. higher chance of being very far away from the mean).

REFERENCES

Daron Acemoglu and Joshua Angrist. How large are the social returns to education? evidence from compulsory schooling laws. Technical report, National bureau of economic research, 1999.

Joshua D Angrist and Alan B Krueger. Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics, 106(4):979–1014, 1991.

Joshua D Angrist and Alan B Krueger. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. Journal of the American statistical Association, 87(418):328–336, 1992.

McKinley L Blackburn and David Neumark. Omitted-ability bias and the increase in the return to schooling. Journal of labor economics, 11(3):521–544, 1993.

Adrian Colin Cameron and Pravin K Trivedi. Microeconometrics using Stata, volume 2. Stata press College Station, TX, 2010.

David Card. Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research, 1993.

David Card. Estimating the return to schooling: Progress on some persistent econometric problems. Econometrica, 69(5):1127–1160, 2001.

Charlotte Christiansen, Juanna Schröter Joensen, and Helena Skyt Nielsen. The risk-return trade-off in human capital investment. Labour Economics, 14(6):971–986, 2007.

Jan S Cramer, Joop Hartog, Nicole Jonker, and C Mirjam Van Praag. Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. Journal of economic behavior & organization, 48(1):29–36, 2002.

German Cubas and Pedro Silos. Career choice and the risk premium in the labor market. Review of Economic Dynamics, 26:1–18, 2017.

Flavio Cunha and James J Heckman. The evolution of inequality, heterogeneity and uncertainty in labor earnings in the us economy. Technical report, National Bureau of Economic Research, 2007.

Serge Darolles, Yanqin Fan, Jean-Pierre Florens, and Eric Renault. Nonparametric instrumental regression. Econometrica, 79(5):1541–1565, 2011.

Louis Eeckhoudt and Harris Schlesinger. Changes in risk and the demand for saving. Journal of Monetary Economics, 55(7):1329–1336, 2008.

Paul HC Eilers and Brian D Marx. Flexible smoothing with b-splines and penalties. Statistical Science, pages 89–102, 1996.

Paul HC Eilers and Brian D Marx. Splines, knots, and penalties. Wiley Interdisciplinary Reviews: Computational Statistics, 2(6):637–653, 2010.

Jesper Ekelund, Edvard Johansson, Marjo-Riitta Järvelin, and Dirk Lichtermann. Self-employment and risk aversionevidence from psychological test data. Labour Economics, 12(5):649–659, 2005.

Tazeen Fasih, Geeta Kingdon, Harry Anthony Patrinos, Chris Sakellariou, and Mans Soderbom. Heterogeneous returns to education in the labor market. The World Bank, 2012.

Peter Hall and Jeffrey S Racine. Infinite-order cross-validated local polynomial regression. Journal of Econometrics, 185:510–525, 2015.

Colm Harmon and Ian Walker. Estimates of the economic return to schooling for the united kingdom. The American Economic Review, 85(5):1278–1286, 1995.

Colm Harmon, Vincent Hogan, and Ian Walker. Dispersion in the economic return to schooling. Labour economics, 10(2):205–214, 2003a.

Colm Harmon, Hessel Oosterbeek, and Ian Walker. The returns to education: Microeconomics. Journal of economic surveys, 17(2):115–156, 2003b.

Joop Hartog and Wim PM Vijverberg. On compensation for risk aversion and skewness affection in wages. Labour Economics, 14(6):938–956, 2007.

Tristen Hayfield and Jeffrey S Racine. Nonparametric econometrics: The np package. Journal of Statistical Software, 27:1–32, 2008.

James J Heckman, Lance J Lochner, and Petra E Todd. Earnings functions, rates of return and treatment effects: The mincer equation and beyond. Handbook of the Economics of Education, 1:307–458, 2006.

James J Heckman, Lance J Lochner, and Petra E Todd. Earnings functions and rates of return. Journal of human capital, 2(1):1–31, 2008.

Daniel J Henderson and Christopher F Parmeter. Applied Nonparametric Econometrics. Cambridge University Press, 2015.

Daniel J. Henderson, Solomon W. Polachek, and Le Wang. Heterogeneity in schooling rates of return. Economics of Education Review, 30:1202–1214, 2011a.

Daniel J Henderson, Solomon W Polachek, and Le Wang. Heterogeneity in schooling rates of return. Economics of Education Review, 30(6):1202–1214, 2011b.

Lennart Hoogerheide, Joern H Block, and Roy Thurik. Family background variables as instruments for education in income regressions: A bayesian analysis. Economics of Education Review, 31(5):515–523, 2012.

Joel L Horowitz. Applied nonparametric instrumental variables estimation. Econometrica, 79(2):347–394, 2011.

Michael F Hutchinson and FR De Hoog. Smoothing noisy data with spline functions. Numerische Mathematik, 47(1):99–106, 1985.

Barış Kaymak. Ability bias and the rising education premium in the united states: A cohort-based analysis. Journal of Human Capital, 3(3):224–267, 2009.

Miles Kimball and Philippe Weil. Precautionary saving and consumption smoothing across time and possibilities. Technical report, National Bureau of Economic Research, 1992.

Miles S Kimball. Precautionary saving in the small and in the large. Econometrica: Journal of the Econometric Society, pages 53–73, 1990.

Gary Koop and Justin L Tobias. Learning about heterogeneity in returns to schooling. Journal of Applied Econometrics, 19(7):827–849, 2004.

Simon Kuznets and Milton Friedman. Incomes from independent professional practice, 1929-1936. In Incomes from Independent Professional Practice, 1929-1936. NBER, 1939.

Shujie Ma, Jeffrey S Racine, and Lijian Yang. Spline regression in the presence of categorical predictors. Journal of Applied Econometrics, 30:705–717, 2015a.

Shujie Ma, Jeffrey S Racine, and Lijian Yang. Spline regression in the presence of categorical predictors. Journal of Applied Econometrics, 30(5):705–717, 2015b.

Pedro S Martins and Pedro T Pereira. Does education reduce wage inequality? quantile regression evidence from 16 countries. Labour economics, 11(3):355–371, 2004.

Jacopo Mazza, Hans van Ophem, and Joop Hartog. Unobserved heterogeneity and risk in wage variance: Does more schooling reduce earnings risk? Labour Economics, 24: 323–338, 2013.

Kimmarie McGoldrick and John Robst. The effect of worker mobility on compensating wages for earnings risk. Applied Economics, 28(2):221–232, 1996.

Costas Meghir and Mårten Palme. Educational reform, ability, and family background. American Economic Review, 95(1):414–424, 2005.

Eduardo L Montoya, Nehemias Ulloa, and Victoria Miller. A simulation study comparing knot selection methods with equally spaced knots in a penalized regression spline. International Journal of Statistics and Probability, 3(3):96, 2014.

Kevin M Murphy and Finis Welch. Empirical age-earnings profiles. Journal of Labor economics, 8(2):202–229, 1990.

Elizbar A Nadaraya. On estimating regression. Theory of Probability & Its Applications, 9(1):141–142, 1964.

Whitney K Newey, James L Powell, and Francis Vella. Nonparametric estimation of triangular simultaneous equations models. Econometrica, 67(3):565–603, 1999.

Zhenghua Nie and Jeffrey S Racine. The crs package: Nonparametric regression splines for continuous and categorical predictors. R Journal, 4(2), 2012.

Deniz Ozabaci, Daniel J Henderson, and Liangjun Su. Additive nonparametric regression in the presence of endogenous regressors. Journal of Business & Economic Statistics, 32(4):555–575, 2014.

Ignacio Palacios-Huerta. An empirical analysis of the risk properties of human capital returns. American Economic Review, 93(3):948–964, 2003.

Simon C Parker and C Mirjam Van Praag. Schooling, capital constraints, and entrepreneurial performance: The endogenous triangle. Journal of Business & Economic Statistics, 24(4):416–431, 2006.

Jeffrey S Racine, Zhenghua Nie, Brian D Ripley, and Maintainer Jeffrey S Racine. Package crs. 2018.

David Ruppert, Matt P Wand, and Raymond J Carroll. Semiparametric Regression. Number 12. Cambridge University Press, 2003.

Liangjun Su and Aman Ullah. Local polynomial estimation of nonparametric simultaneous equations models. Journal of Econometrics, 144:193–218, 2008.

Simone N Tuor and Uschi Backes-Gellner. Risk-return trade-offs to different educational paths: vocational, academic and mixed. International journal of Manpower, 31 (5):495–519, 2010.

Geoffrey S Watson. Smooth regression analysis. Sankhyā: The Indian Journal of Statistics, Series A, pages 359–372, 1964.

John V Winters. Estimating the returns to schooling using cohort-level maternal education as an instrument. Economics Letters, 126:25–27, 2015.