Article

Higher-Order Risk–Returns to Education

Daniel J. Henderson [1,2,3,*], Anne-Charlotte Souto [4] and Le Wang [5]

[1]   Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL 35487, USA
[2]   School of Mathematical Sciences, Nankai University, Tianjin 300071, China
[3]   Institute for the Study of Labor (IZA), 53113 Bonn, Germany
[4]   Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260, USA; acharlotte.souto@pitt.edu
[5]   Department of Economics, University of Oklahoma, Norman, OK 73019, USA; le.wang.econ@gmail.com
[*]   Correspondence: djhender@cba.ua.edu

check for
updates

Abstract: In the traditional human capital framework, education is often considered as an investment, rather than consumption, while consumption is not necessarily precluded. Whether education is an investment is empirically unclear and relatively under-explored. We shed light on this issue by estimating the risk–return trade-off in the context of education. If education is indeed an investment, risk could play an important role in individual educational decisions just as with risky assets. As portfolio theory predicts, there could be a trade-off between returns to education and risks concerning those returns: higher risks are generally associated with higher returns. We contribute to the literature by proposing various measures of risk based on the entire distribution of returns to education recovered by our nonparametric models. Our results confirm a trade-off between returns and variance. We also found statistically significant impacts for the higher moments: skewness and kurtosis. Interestingly, we found the relationship between mean returns and variance to be linear, and the relationship between expected returns and higher-moments (skewness and kurtosis) is non-linear.

Keywords: return to education; risk; higher moments; portfolio theory; categorical regression splines

1. Introduction

Education has been considered both as an investment and consumption good (Stiglitz 1974; Schaafsma 1976). The investment aspect of education is similar to financial assets in that they both affect future income and well-being, and investments in human capital and financial assets are both associated with considerable uncertainty (although insurance options for reducing riskiness in human capital accumulation do not exist—Shaw 1996). To that end, 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). If individuals consider education to be an investment good, as opposed to a consumption good, decisions about the level of investment in education should be made in the same way as the decisions about any other risky assets.

Extensive theoretical and empirical literature exist linking an individual's decision making under uncertainty with their risk attitudes, which are the basis of portfolio theory for risky assets. Most of the focus has been on second-order risk, known as risk aversion (starting with Bernoulli 1738). Risk averse investors prefer a less risky portfolio to a riskier one for a given mean return. This assumption implies that an investor will take on more risk only if he/she is expecting higher returns. Empirically, we should observe a positive correlation between the risk-level of a portfolio, usually measured by the variance (second-order risk) of the distribution of its returns, and its expected return (i.e., the mean of the distribution of its returns); the so-called risk–return trade-off. Among the relatively scarce literature examining the investment aspect of education, emphasis has been placed on testing the risk–return trade-off, for which empirical evidence supporting it has been found.

The traditional portfolio, however, has neglected the role of higher-order risk attitudes in describing individuals' risk preferences. Samuelson (1970) argued that only including second-order risk implies restrictions on the distribution of outcomes. 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 relate to an 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). These recent theoretical developments imply that, in a risky investment decision context, individuals would require a premium (via a higher mean return) not just when facing second-order risk (i.e., high variance) but also when facing higher-order risks, such as third and fourth-order risk (i.e., negative skewness and high kurtosis). Moreover, it is important to control for these different types of risks simultaneously when examining the risk–return trade-off unless we believe they are not related with one another.

We go beyond second-order risks to examine the relationship between rates of return to education and higher-order risks. We ask the following questions: Are the risks in returns compensated by higher mean returns (i.e., is education an investment good)? Is the distribution's variance sufficient to capture the risk taken by individuals when they are making education investment decisions? Is any type of risk itself an omitted variable when studying the risk–return trade-off for human capital investment?

Our approach differs from the prior literature in that we are able to estimate individual-specific returns to education and hence the *entire* distribution of the returns by using nonparametric methods. This approach allows us to capture additional features of the distribution of returns to education (including the mean, variance, skewness and kurtosis, as our main variables of interest). Our approach also allows for testing the importance of each order of risk simultaneously. Using data from the March Current Population Survey (CPS), we found that risk, as measured by higher-moments of the rate of return distributions, indeed plays an important role in the educational decisions of individuals, and that we should consider education as an investment good.

Similar to the impact in the literature on log wages, we found a trade-off between mean returns to education and the variance. We simultaneously studied the trade-off between mean returns to education and higher-order risks.[1] We found that 60% and 80% of the compensations for skewness and kurtosis, respectively, are statistically significant. Importantly, we found the relationship between mean returns and our higher-moments to be nonlinear. Overall, our empirical results suggest that the relationship between educational risk and returns requires more than the second-moment of the distribution and that this relationship cannot be sufficiently captured via linear parametric models.

The remainder of our paper is organized as follows: Section 2 gives a literature review and highlights our main contributions. The Section 3 presents our methodology, while the fourth describes our data. In Section 5, we present and analyze our empirical results. Section 6 concludes.

2. Literature Review and Our Contributions

Our work is related to the earnings dispersion literature. Kuznets and Friedman (1939) first discovered a positive relationship between the mean and variance of incomes among workers (from a selected few professions). In subsequent work, authors 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 fields and/or types of education

---

[1] See Hartog and Vijverberg (2007) for a study on second and third-order effects of risk with respect to log wages in a parametric setting.

(Christiansen et al. 2007; Backes-Gellner et al. 2010) and across different levels of general education (Palacios-Huerta 2003). By including risk preferences (such as second-order risk attitude) into a basic model, researchers have been able to better predict individuals' choices in the presence of risk/uncertainty (Ghosh and Ray 1997).

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 educational context uses either moments from the distributions of estimated earnings or 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, obtain the higher moments (variance and occasionally 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 earnings dispersion, we estimate 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 as captured by earnings (or residuals) dispersion.

Second, in addition to the variance (second moment), we also use the third and fourth moments (skewness and kurtosis) of the rate of returns' distribution to capture risk. Only a few papers have looked at the relationship between higher-order risk, particularly skewness, and the mean, and none so far (as we are aware of) has examined the trade-off between kurtosis (fourth-order risk) and mean returns. The theoretical literature on decision making under uncertainty suggests that higher-order risk attitudes also matter when individuals make decisions. By adding those higher moments into our model, we can investigate if there also exists compensation for the third and fourth moment risk-types. If we find compensation for those other types of risk, it would suggest that the variance does not fully capture the risk of investing in education and that higher-order risk attitudes play a role in education investment decisions.

Third, as opposed to the parametric structure (log-linear model) that the literature has typically imposed, we model the relationship between mean returns and higher-order risks nonparametrically. That is, we allow the relationship between years of schooling and earnings and the relationship between risk of investing in one more year of schooling and expected earning to be nonlinear. The relationship between risk and expected earnings is unknown and likely to be nonlinear. For example, compensation for risk may vary with the level of expected returns. Also, risk attitudes are likely to be heterogeneous. This cannot be captured by a single, constant parameter without assumptions on the underlying preferences of individuals (e.g., constant relative risk aversion). For example, risk attitudes might have changed over time (i.e., individuals in the 1980s might be more risk-averse than individuals in the 2010s) and would require more compensation for risk (i.e., have different gradients). Our nonparametric model allows us to observe this potential heterogeneity which so far has been ignored in the literature.

3. Methodology

To estimate if there are risk–return trade-offs to education, we employed a four step procedure. Our approach was as follows: (1) estimate an individuals' rate of return to education; (2) merge those rates into 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 a Mincer-type equation which regresses log wage on education and experience. More formally, we nonparametrically regress the log of an individuals' weekly wage on education and other control variables for each of the 8 years in our sample separately via

$$ln(wage_{it}) = m(educ_{it}, X_{it}) + \epsilon_{it}, \tag{1}$$

for individual i in each year $t = 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015$, where *wage* is the weekly wage, *educ* corresponds to the number of years of education (beginning with first grade) and X is a vector of control variables (e.g., years of experience, marital status, metropolitan area, etc.).

As our sample includes both continuous and discrete variables, we will use categorical regression splines (CRS), a method developed by Ma et al. (2015), to deal with categorical variables that cannot be accommodated by traditional spline regression. In short, the method combines spline regression (for continuous variables) and discrete kernel weights (for categorical variables). In our case, continuous variables such as education and experience will be modeled using splines, while categorical variables (such as race and metropolitan area) will be incorporated using kernel weights.[2]

As extensively discussed in the literature (Card 1999; Card 2001), schooling choice may be related to unobservable variables such as innate ability, which can bias our estimates (i.e., omitted variable bias). To address this issue, we use the nonparametric instrumental variable (IV) approach developed in Horowitz (2011).[3] This approach is desirable relative to alternative IV approaches in the literature such as Darolles et al. (2011)'s approach, given the relative computational requirements.

Horowitz (2011)'s approach is analogous to the IV approach in the linear case (OLS with IV) and thus has similar requirements for its instruments to be valid. Specifically, this approach assumes existence of an IV, W, such that for Equation (1):

$$E(\epsilon | W = W_{it}, X = X_{it}) = 0, \tag{2}$$

where W is an instrument for education. It follows that, using the CRS methodology, we estimate the unknown function $m(\cdot)$ such that

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

Similar to parametric IV methods, we want an instrument W that is correlated with years of education (*educ*), but not with ability (and thus will not affect weekly wage directly). Unlike the parametric IV approach, the number of instruments cannot exceed the number of endogenous variables. In our case, as we consider education to be the only endogenous variable in our model, we will only use a single instrument (discussed in the data section).

In step 2, we obtain individual-specific returns to education using nonparametric estimation. Specifically, for each of our $t = 1, 2, \ldots, T$ (8 years) regressions,[4] we obtain gradients for each individual as

$$rr_{it} = \frac{\partial ln(wage_{it})}{\partial educ_{it}}. \tag{3}$$

In step 3, we compute both average returns and various measures of risk for each state j, in each time period t. Specifically, we obtain the first four moments for our analysis of the risk–return trade-off in education as

---

2   According to Ma et al. (2015)'s simulations results, for a small number of continuous variables (less than 5) and a relatively large sample size (more than 500), as we have here, CRS results in 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 (which is only consistent if the model is additively separable).

3   Specifically, we use the CRS R package developed by Nie and Racine (2012) (recently updated: Racine et al. (2018)).

4   Note that all of our model parameters (i.e., our choice of degree, segments (number of knots) and knots' location for splines and of lambda (bandwidths) for our kernels), are selected via cross-validation.

$$
\begin{aligned}
Mean_{tj} &= \overline{rr}_{tj} \\
Var_{tj} &= \frac{1}{N_{tj}} \sum_i (rr_{itj} - \overline{rr}_{tj})^2 \\
Skew_{tj} &= \frac{1}{N_{tj}} \sum_i (rr_{itj} - \overline{rr}_{tj})^3 \\
Kurt_{tj} &= \frac{1}{N_{tj}} \sum_i (rr_{itj} - \overline{rr}_{tj})^4,
\end{aligned}
$$

where $\overline{rr}_{tj}$ is the sample average (over individuals i) for state j in year t; "*Var*" is the variance; "Skew", the skewness, measures the degree and direction of asymmetry in the distribution; and "Kurt", the kurtosis, measures the heaviness of the tails. N_{tj} is the sample size for each of the 408 subsamples (=51 states in each of the 8 years).[5]

In our final step, we use a standard spline regression to regress the mean of returns to education[6] on its corresponding higher moments (our measure of risk) for each subsample:

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

where our nonparametric model allows the relationship between the measure of risk and the return to education to be potentially nonlinear. There are several reasons why such nonlinearity may exist. First, a risk-averse individual may require a different level of compensation at a different level of risk, income and/or education level. For example, an 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. Further, the trade-off between mean returns and one type of risk may also depend on other types of risks. If the theoretical literature on decision making under uncertainty holds, we expect to see a risk–return to education trade-off, that is, in the fourth step, our gradients on variance and kurtosis will be positive and our gradients on skewness will be negative.

4. 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. Following the literature, we focus our analysis on men[7] in the labor force with a positive income between the age of 20 and 59 (working age) living in a household (excluding group quarters). We also exclude self-employed men as their income measures tend to be less reliable and they tend to have different risk tolerances than typical wage earners (Cramer et al. 2002; Ekelund et al. 2005). Our final potential sample was composed of 245,401 men with information on their 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 dummies).

---

5   We are abusing the definition and calling the District of Columbia a state here.
6   Most papers in the literature (e.g., Hartog and Vijverberg 2007) parametrically regress log wage on risk, obtained via the residuals of a first stage Mincer regression, whereas we are using the mean returns themselves.
7   The literature has typically focused on men as women's labor force participation rates are generally low and there exists sample selection for working women (Henderson et al. 2011). Not only are the nonparametric methods to address this issue challenging (Ozabaci and Henderson 2015), but they often require an exclusion restriction that can be difficult to find, let alone be non-controversial. Given that our main contribution is the methodology, we leave the analysis of women to future research.

Following the literature on the estimation of the Mincer earnings equation, we use log weekly wages as our dependent variable,[8] after adjusting for inflation. We exclude 69 individuals with zero weekly wages. In our model, we also control for a dummy variable indicating whether or not wages are top-coded (one if yes, zero otherwise).[9] Education, our main independent variable of interest, is simply the number of years spent in school, starting from first grade.

When estimating the returns to education, we control for the following variables. As we do not have information about individuals' work experience, we proxy it using 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. The age group (ordered) variable *age* 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 1 if the individual is not white and zero otherwise. We control for several household characteristics: children and marital status. The *child* dummy is equal to 1 if any children are living in the same house. The *married* dummy is equal to 1 if the individual is both married and living with his spouse, and equal to 0 otherwise (e.g., the individual is married but the spouse is absent, or if the individual is divorced, widowed, or has never been married). Finally, we control for regional effects using nine regional dummies and for city effects using a metropolitan (*metro*) dummy which equals 1 if the individual lives in a central city or right outside a central city.

For the nonparametric IV approach, we borrow a measure of family characteristics. The literature has proposed some candidates for an IV, but all of them have been shown to have some limitations. For example, some used 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), a combination of birth and reforms (Acemoglu and Angrist 1999), a 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 spouse's yearly wage, a measure of family characteristics, as our IV for two reasons. First, it is a good proxy for a spouse's productivity and ability. Considering the theory of positive assortative matching on education in marriages, this would be highly related to one's education, but not necessarily directly related to one's wages. Second, while our data include parents' education, spouse's education, parents' income and spouse's income (bottom of Table 1), this measure minimizes the loss of information due to missing values.[10] We need to emphasize that our goal here is not to debate about the validity of these IVs, but rather use one to illustrate our proposal in evaluating the risk–return trade-off. For alternative IVs, one can readily employ our approach to assess the robustness of our results.

In our third step, we use our state variables to obtain mean returns and various risk measures for state local markets. That is, we infer that each state's labor market faces different risks due to different "local" conditions and laws.

From the descriptive analysis of our full sample (Table 1), we can 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; 13.8% of our sample was non-white, 67.5% was married (with the spouse present in the household) and 55.7% had at least one child living with them. Note that the last five variables in Table 1 are the five possible instrumental 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. On average, the spouse of the men

---

[8] We use weekly wage instead of yearly wage to control for productivity.
[9] Data that is top-coded reduces the chance of a right-skewed distribution of wages.
[10] This can create further sample selection issue. In the Chinese context, Wang (2013) shows that the issue is potentially more severe when using parental characteristics than spousal characteristics.

in our sample earned $17,175 per year (about $488 per week). We also considered using the spouse's weekly wage, but favored yearly wage as it was reported at a higher rate.[11,12]

Table 1. Descriptive statistics.

Statistic	*n*	Mean	St. Dev.
survey year	245,401	1998	11.820
occupation	245,401	14.838	8.010
log weekly wage	245,332	6.474	0.732
weekly wage	245,401	830.469	744.062
education	245,401	13.393	2.867
experience	245,401	19.153	10.798
married	245,401	0.675	0.468
age	245,401	38.545	10.584
age_20's	245,401	0.242	0.428
age_30's	245,401	0.301	0.459
age_40's	245,401	0.266	0.442
age_50's	245,401	0.191	0.393
non-white	245,401	0.138	0.345
topcoded	245,401	0.011	0.106
child	245,401	0.557	0.497
metro	245,401	0.603	0.489
new-england	245,401	0.088	0.283
middle atlantic	245,401	0.119	0.324
east-north central	245,401	0.136	0.343
west-north central	245,401	0.100	0.300
south atlantic	245,401	0.165	0.371
east-south central	245,401	0.048	0.213
west-south central	245,401	0.092	0.289
mountain	245,401	0.112	0.315
pacific	245,401	0.142	0.349
father's yearly wage	15,460	38,963.970	34,284.400
spouse's yearly wage	175,498	17,175.330	23,264.220
father's education	15,460	11.973	3.534
mother's education	20,436	11.710	3.325
spouse's weekly wage	131,621	487.934	517.246

5. Results

This section will be broken down according to our four step procedure: (1) we used categorical regression splines with an IV (Ma et al. 2015) to separately estimate the returns to education for each individual (in each time period) in our sample; (2) we merged the data from the first step back into our original dataset; (3) we split the sample by year and state, and obtained the first four moments (mean, variance, skewness, kurtosis) of the rate of returns' distribution for each state in each year; and (4) we regressed the mean (first moment) returns to education on the higher-order moments to see if we would observed compensation for each type of risk.

---

[11] Note that although 165,646 of the men in our sample were married, 175,498 reported their spouses' incomes. This difference is due to men reporting their spouses' incomes while their spouse were not living under the same roof.

[12] In our analysis, we regressed each year separately. While the full sample had between 24,112 and 36,655 individuals per year, we used a random sample of 8000 individuals in each year to improve comparability between years, not to mention improvements in computation time.

5.1. Step 1: Estimated Rates of Returns

In our first step, we regress (using CRS) the log weekly wage on schooling and our control variables (described in the data section) separately for each quinquennial year.[13] The summary statistics of men's gradients on schooling (i.e., returns to schooling) for each year are presented in Table 2.[14] For example, the first row of Table 2 represents the summary statistics of the returns to schooling for all the men in our sample surveyed in 1980. Hence, on average, men surveyed in 1980 earned an additional 7.37% on their weekly wage for one more year spent in school. Similarly, the median man earned an additional 7.53%.

Table 2. Return to education (Step 1) summary statistics.

	1st Qu.	Median	Mean	3rd Qu.
1980	0.0666	0.0753	0.0737	0.0866
1985	0.0755	0.0923	0.1011	0.1273
1990	0.1544	0.2231	0.2162	0.2904
1995	0.1439	0.1791	0.1887	0.2364
2000	0.2006	0.2449	0.2465	0.2995
2005	0.2493	0.2807	0.2716	0.3078
2010	0.1340	0.1849	0.1753	0.2248
2015	0.1488	0.2299	0.2217	0.2925

For completeness, we also performed parametric regressions (adding polynomial terms for education and experience) with and without an IV (spouse's wage). We found the estimated rate of returns range from about 6% to 12% in the models without IVs, and from about 4% to 23% in the models with IVs. These results are in line with what is commonly found in the literature (Harmon et al. 2003). While the estimated rate of returns are similar, we do see important differences, especially in terms of our ability to handle dispersion (which we highlight below).

Overall, we note a general increase in individuals' returns to education over time. This general trend is typically found in the literature (Harmon et al. 2003; Deschênes 2006). The two panels in Figure 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 how the range of returns to education varied substantially across states and 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% upper and lower bounds of the the estimates on the gradients on schooling.[15] Apart from a general increase throughout the whole period, we note a significant drop in mean returns in 2010, the aftermath of the financial crisis, followed by a rebound in 2015.

---

[13] All regressions had knots placed at equally spaced quantiles. The number of degree and segments (i.e., knots minus one) were all selected via cross-validation (CV) with 8 restarts from different random initial points (5 is the default, we increased it to a higher number to avoid local optima). The number of segments selected via CV for both schooling and experience were between 1 and 5 and the number of degrees were either 2 or 3. We used the default regularization method "Landweber-Fridman."

[14] The R-squared for each of those regressions ranged from 0.30 to 0.65.

[15] The upper and lower estimates are often mistaken as an alternative to the commonly used confidence bounds in parametric models.
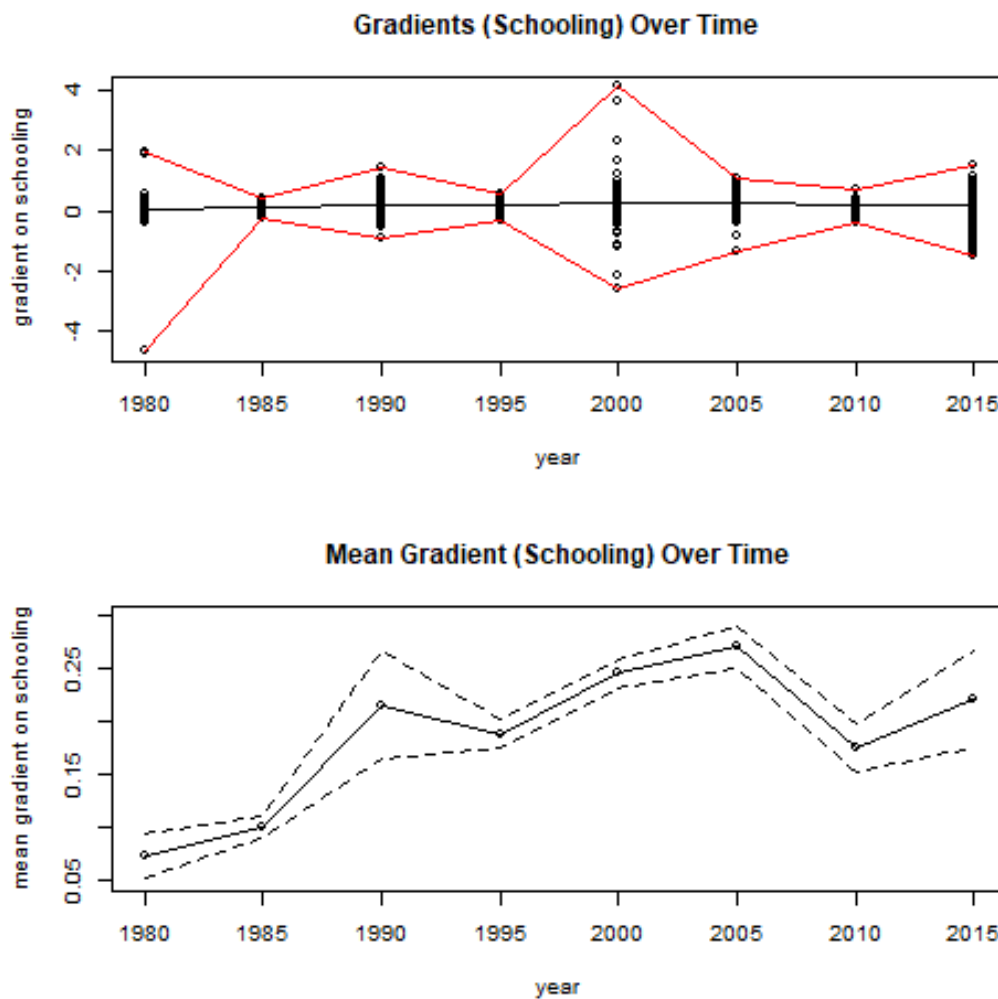
Figure 1. Gradients on schooling (education) from Step 1.

Figure 2 shows the resulting density of rate of returns to education for each year. Note that those distributions do not appear to be Gaussian. Other than perhaps for 1980, they all appear to be skewed (to varying degree). Their variance and tails (kurtosis) also seem to vary from year to year. The results here are similar to those in Henderson et al. (2011), who study the distribution of returns to education over time, but we note that they did not control for endogeneity in their study.

5.2. Steps 2 and 3: Moments by Year and State

In our second step, using these individual-specific gradients, we can compute mean returns and various risk measures for each state labor market in each year.[16] We compute four moments for each state market in each year: mean, variance, skewness and kurtosis. We thus have 408 observations (=51 states in each of our 8 years) for each moment. Table 3 presents the summary statistics for each of the 4 moments obtained. To have a better idea of what the data look like, we also plot the density of each moment in Figure 3.

---

[16] The plots of those distributions are available upon request.

Table 3. Summary statistics of moments.

	1st Qu.	Median	Mean	3rd Qu.
mean	0.14	0.20	0.19	0.24
variance	0.03	0.05	0.05	0.07
skewness	0.00	0.01	0.02	0.02
kurtosis	0.00	0.00	0.02	0.01
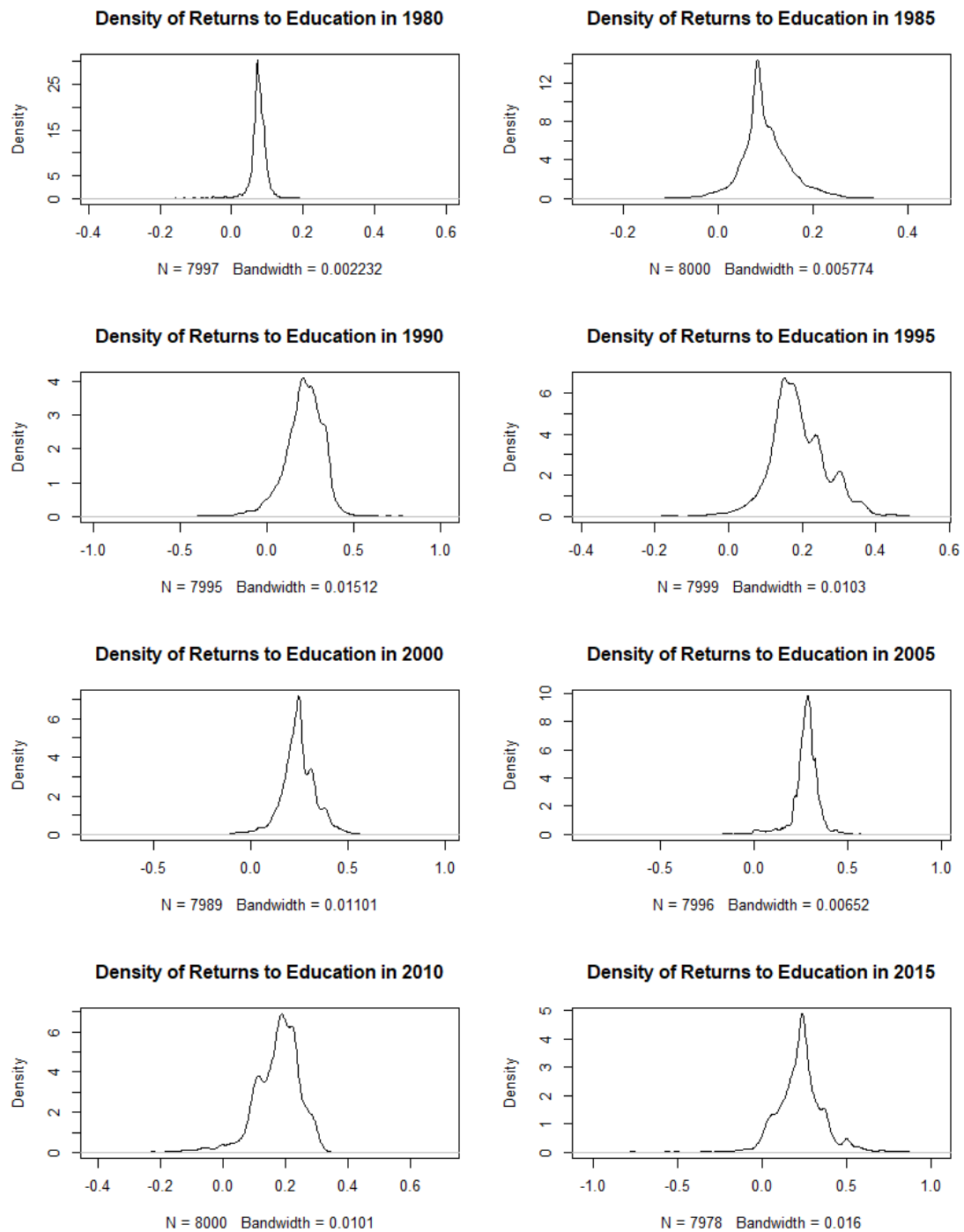
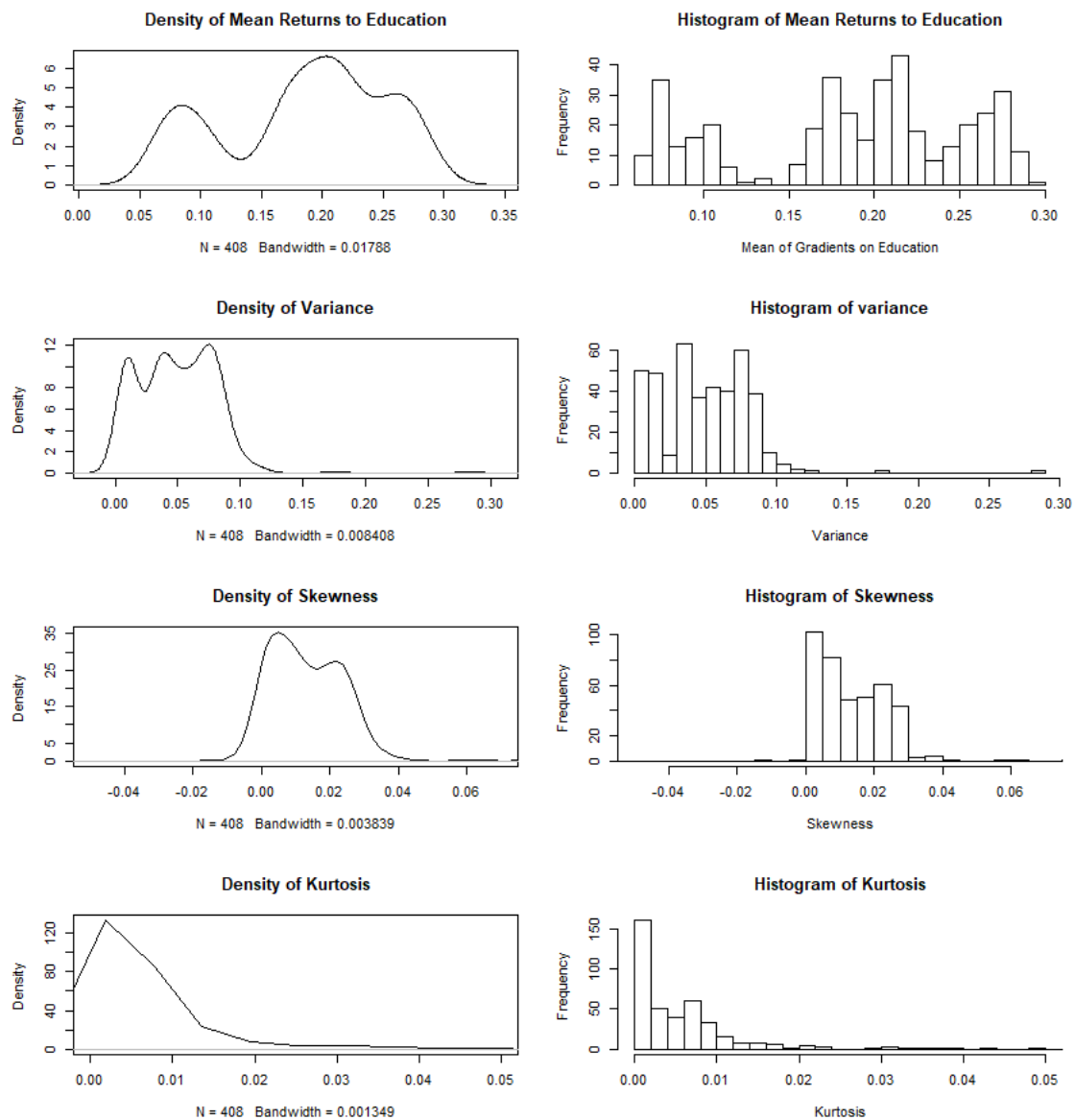**Figure 2.** Densities of gradients on schooling (education) from Step 1.

Figure 3. Moments density.

The density of mean returns appears to have two modes (one near 0.08 and another near 0.20). On the other hand, the densities for the other three moments appear flat or unimodal. The variance typically goes from 0 to 0.10, while the skewness has some negative cases (not uncommon, see Hartog and Vijverberg 2007), but most lie between 0 and 0.04. Kurtosis is also relatively minor with values that lie between 0 and 0.02.

5.3. Dynamics of Each Moment

To get a better idea of the variation that exists in the data, Figures 4–7 show how each moment varies over time with each circle representing one state-year value. Each figure highlights the top 5 (top panel) and bottom 5 (bottom panel) states (in terms of magnitude) over the sample period. For example, the District of Columbia had on average the highest expected returns over the period (Figure 4: top panel) while New Mexico had the lowest expected returns (Figure 4: bottom panel).

Although expected returns have been generally increasing over time, Figure 4 shows that they decreased in both the period between 1990 and 1995 and the period between 2005 and 2010. Such decreases could be attributed to the small recession of 1990 and the economic crisis of 2008, respectively.[17] The range of expected returns were particularly large in 1995 and 2000.

We also note that variance is positively correlated with time (year), and more specifically, it has generally been rising over time. The correlation between the variance and year is roughly 0.60. This is different from the results of Harmon et al. (2003) 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 did not increase over time. They link this result to the rapid increase 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. Our results are able to support that theory.

Figure 5 shows a general increase in the variance over time for both the top and bottom five states. Note that a high variance indicates high inequality in rate of returns to education within each state. While the year 2000 had unusually high variance 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 6 and 7).

In contrast to the results for mean returns and variance, 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, does not change much over time (a correlation of 0.02 with time). Figures 6 and 7 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 large values.[18] This suggests that the risk faced by individuals varied greatly between states. Note that Minnesota and Kansas stand out again with very large estimates for skewness and kurtosis in 2000.

5.4. An Informal Look at the Risk–Return Trade-Off: Correlation between Mean Returns and Each Type of Risk

Table 4 shows the correlation between each moment. There is a positive correlation between expected returns (mean) and all three higher moments. However, the correlation with the fourth moment (kurtosis) is low and not statistically significant. While the correlation of expected returns with variance and with skewness are both statistically significant, the correlation with skewness is much lower (0.25).

Table 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: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

---

[17] Note that the District of Columbia was an exception in 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.

[18] While 2015 does not have particulary extreme values, it does show a wider range than usual for both skewness and kurtosis.
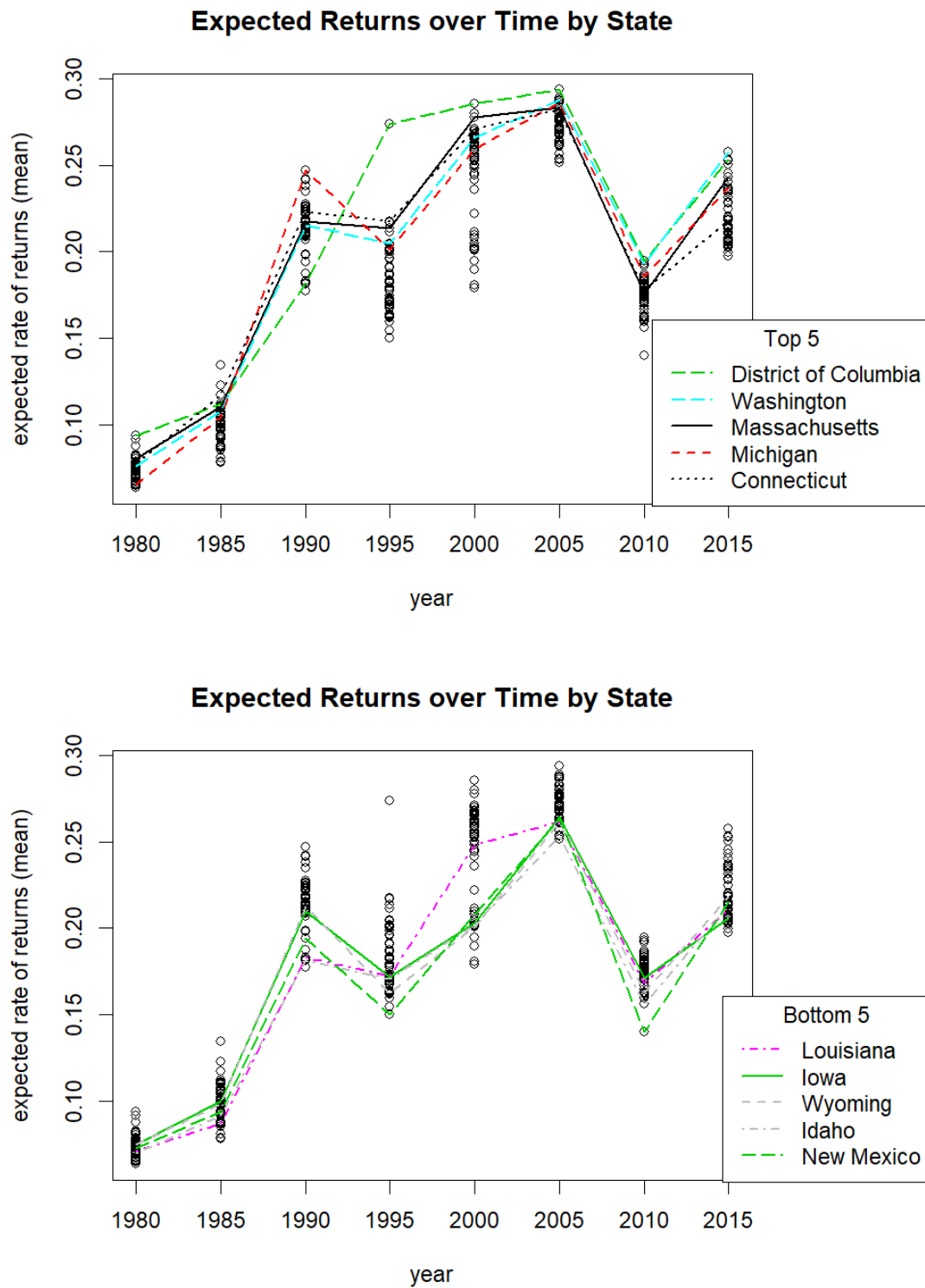
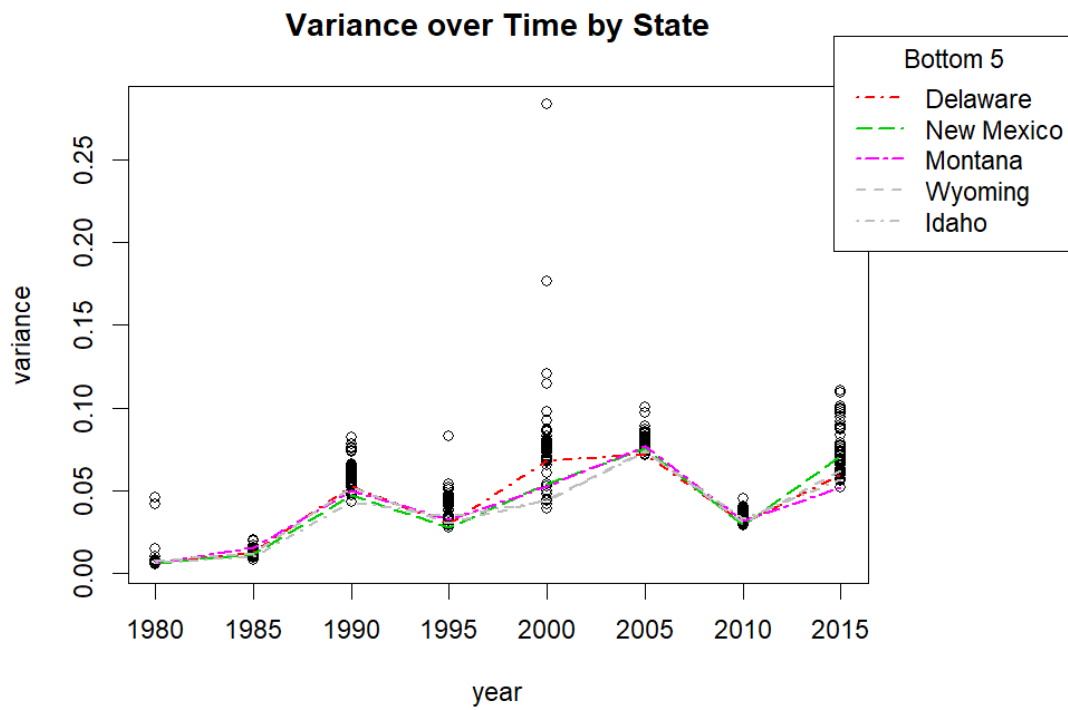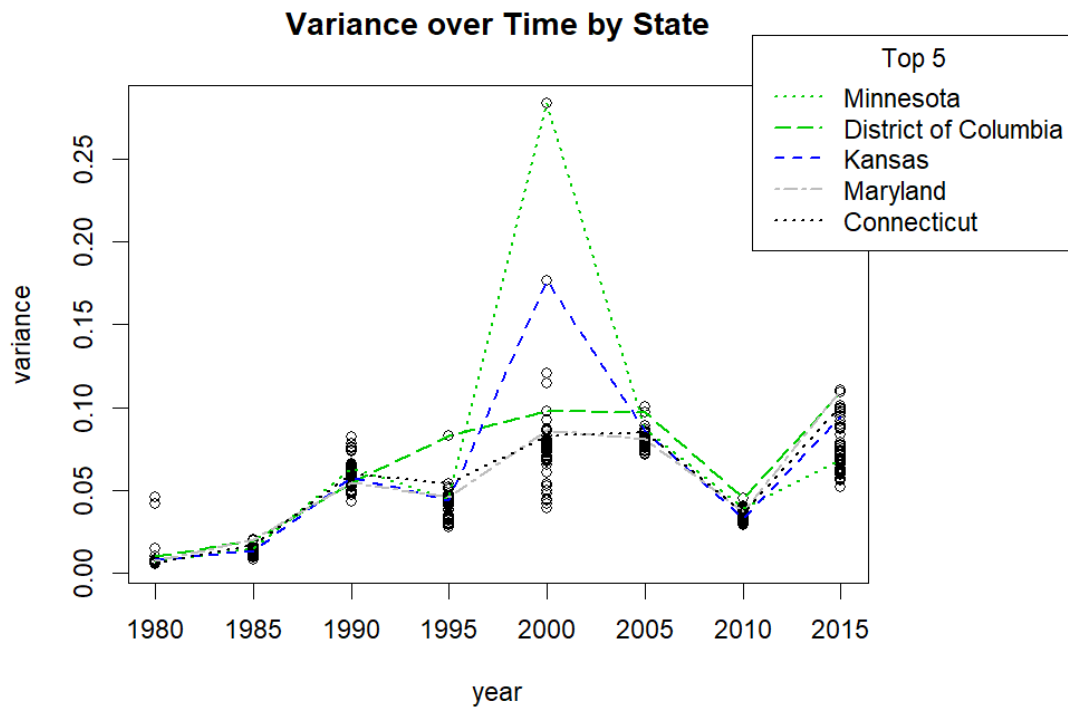**Figure 4.** Expected returns top 5 and bottom 5.

Figure 5. Variance top 5 and bottom 5.

Skewness over Time by State
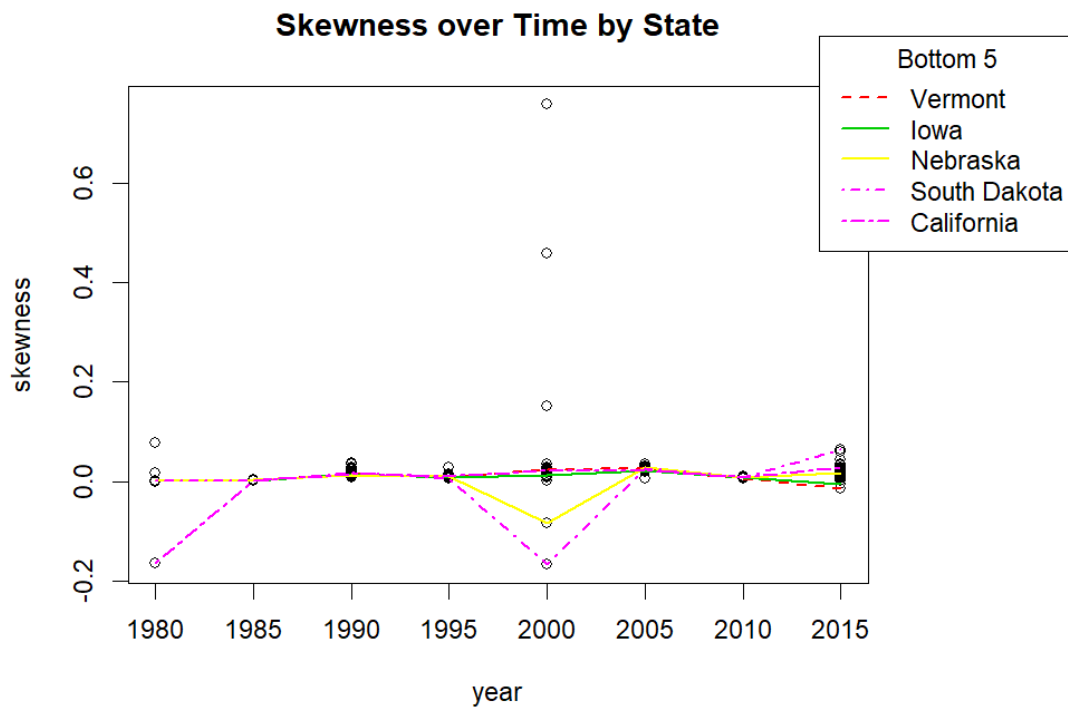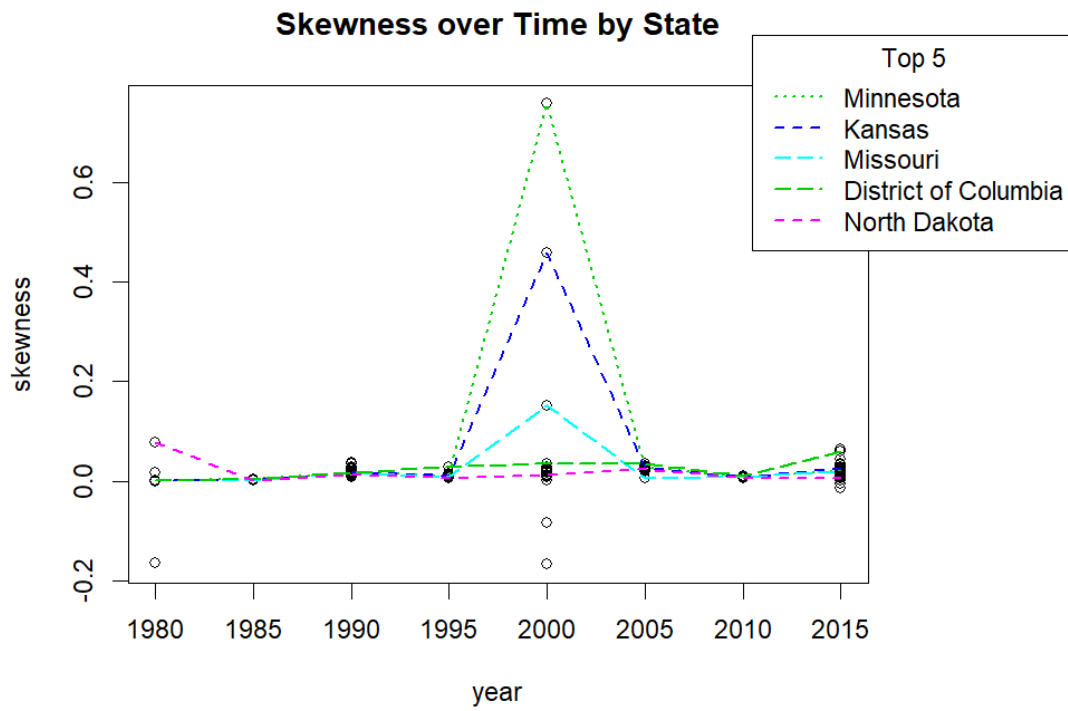

Skewness over Time by State

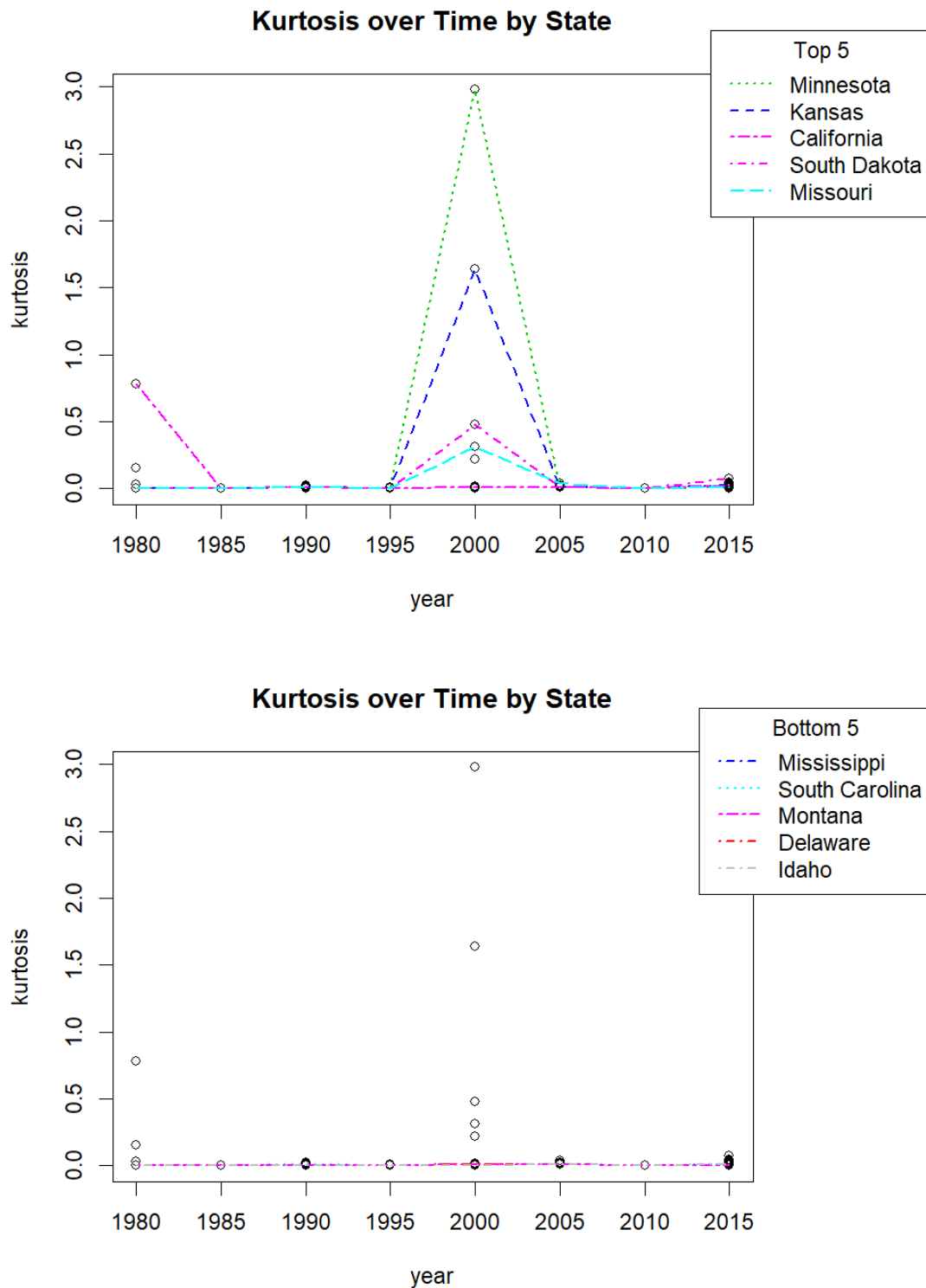**Figure 6.** Skewness top 5 and bottom 5.

Figure 7. Kurtosis top 5 and bottom 5.

Figure 8 plots the higher moments against the expected rate of returns to schooling (mean). For each row of plots, the right panel is a zoom-in of the left panel (i.e., excludes more 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 of returns.
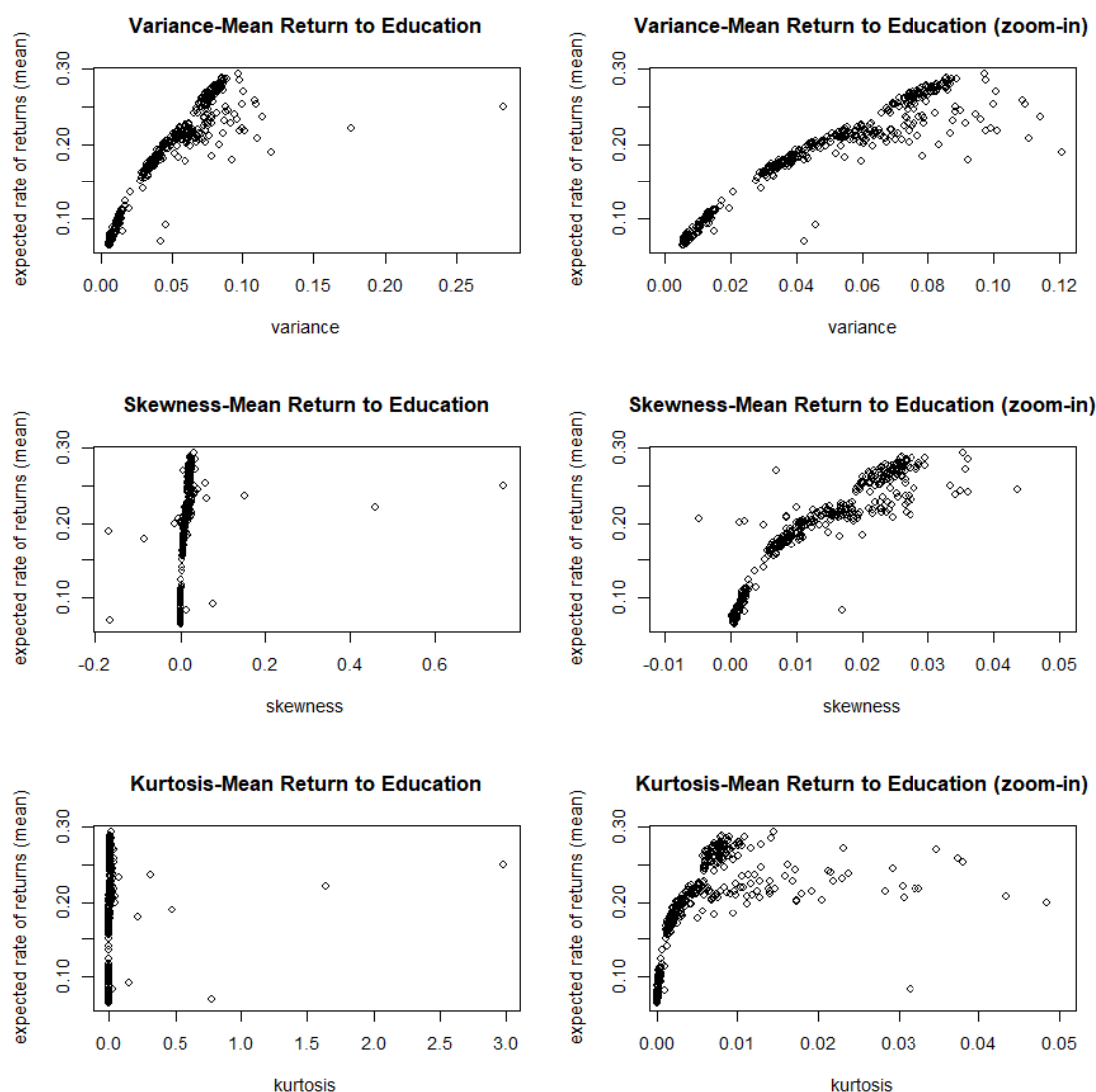
Figure 8. Higher moments against the expected returns (mean).

The second row of plots in Figure 8 show the relationship between skewness and the mean rate of return for 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, the higher the expected returns appear to be. If taken literally, this result indicates that individuals may consider any type of skewness (positive or negative) as a risk and seek compensation for both types. This is in contrast to what is often found in studies regressing log wages on skewness whereby authors argue that workers exhibit skewness affection (Hartog and Vijverberg 2007).

The bottom two panels in Figure 8 plot 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 first and fourth moments of the distributions for very low levels of kurtosis (below 0.01). Looking at the bottom-right panel of Figure 8, we note that the kurtosis values in our sample are all between 0 and 3 (platykurtic distributions). While a few are higher than 0.5, most are very close to zero (i.e., thin tails). This result is interesting as it suggests that there exist relatively few extreme average mean returns in our sample (as evidence by the top and bottom five figures).

5.5. Step 4: Risk–Return Trade-Offs

In our final step (4), we nonparametrically (spline) regress the mean return to education on its higher moments (i.e., variance, skewness and kurtosis) to investigate if there is indeed compensation for risk (while including a time trend and state fixed effects).[19] Our model specification is obtained via cross-validation. The cross-validation procedures determined that the relationship between expected returns and variance is linear, while the relationships between expected returns and the higher moments, skewness and kurtosis, are nonlinear.[20]

Tables 5 and 6 present the summary statistics of the gradients found via this spline regression. In Table 6, we exclude gradients that are not statistically significant. While all of the gradients of variance are constant and statistically significant, about 40% and 20% of the gradients on skewness and kurtosis, respectively, are not statistically significant. As our distributions in our sample have kurtosis values very close to zero (with little variation), it is not surprising that the gradients on those low kurtosis estimates are not statistically different from zero. The gradients on skewness that are not statistically significant are also relatively small values. This results is to be expected as there should not be any additional compensation for symmetric distributions (i.e., skewness close to zero).

Table 5. Spline gradients summary statistics.

	1st Qu.	Median	Mean	3rd Qu.
gradient of variance	1.34	1.34	1.34	1.34
gradient of skewness	−1.38	0.04	0.33	3.04
gradient of kurtosis	−5.35	−1.07	0.74	9.81

Table 6. Spline gradients summary statistics: statistically significant results only.

	1st Qu.	Median	Mean	3rd Qu.	obs.
gradient of variance	1.34	1.34	1.34	1.34	408
gradient of skewness	0.85	3.04	2.05	3.23	245
gradient of kurtosis	−5.35	−1.07	0.23	9.81	326

For comparison purposes, we also show the OLS results (Table 7) 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 first quartile of the skewness gradients 0.85). More strikingly, the OLS coefficient on kurtosis seems to suggest that a higher kurtosis will result in a lower expected return. This contradicts theory. However, the spline results find gradients of far higher magnitude and of both signs, suggesting a far more complex relationship between kurtosis and expected returns.

Table 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

---

[19] See Henderson and Souto (2018) for an intuitive explanation of spline regression with instruments.
[20] The numbers of degrees selected via CV were 1 for variance, skewness and kurtosis, and 2 for the time trend. The numbers of segments (i.e., number of knots minus one) selected via CV were: 1 for variance, 10 for skewness, 10 for kurtosis and 3 for the time trend. The knots were placed at equally spaced quantiles. For the state dummies, a bandwidth of 1 was selected via CV.

While Figure 9 shows the histograms of gradients on skewness and kurtosis, Figures 10–13 plot the gradients of each higher moment against the first moment (left panel) as well as against skewness and kurtosis (right panel). Note that the cross-validation procedure determined that variance enters linearly and so there is no variation (a constant derivative) in the gradients of variance across both the expected returns and the variance itself. Figure 14 shows the statistically significant gradients on 45° plots (Henderson et al. 2012) with their respective 95 percentile confidence bounds.
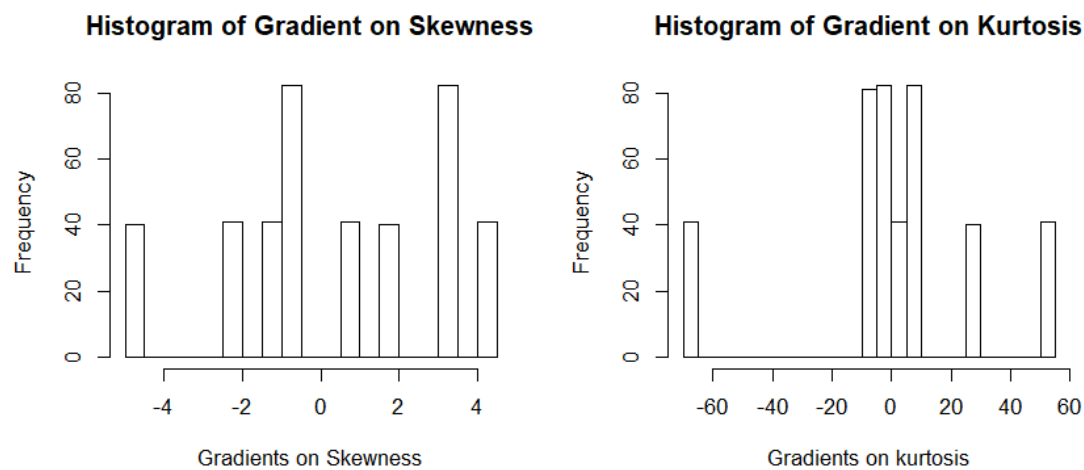
Figure 9. Histograms of spline gradients (Step 4).

The gradients on variance are all positive, as expected, and statistically significant. That is, we do find compensation for second-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 variances 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 units, on average. However, the variance ranges from 0.01 to 0.28. Therefore, an increase of 0.10 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 example, from 20% to 21.34%, is not a negligible amount.
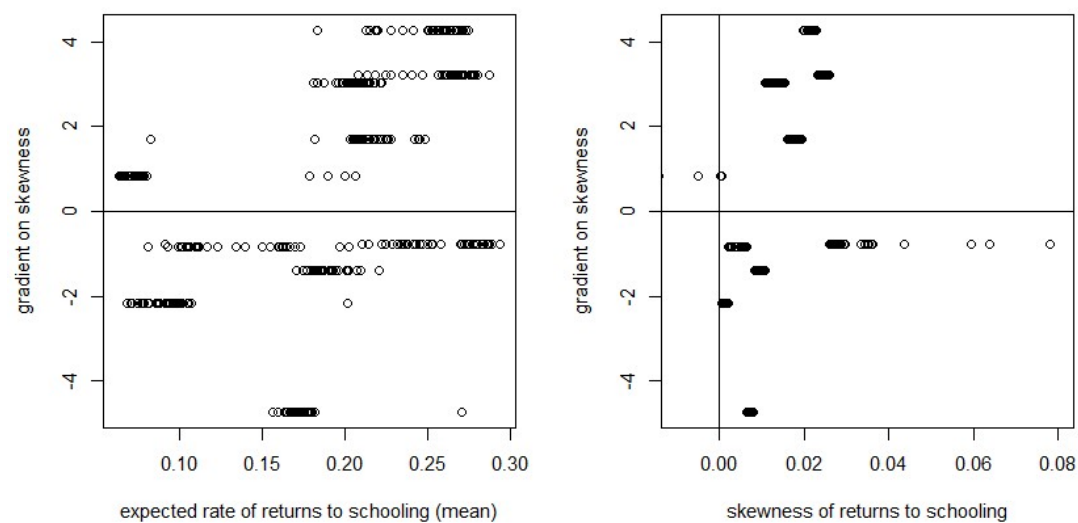

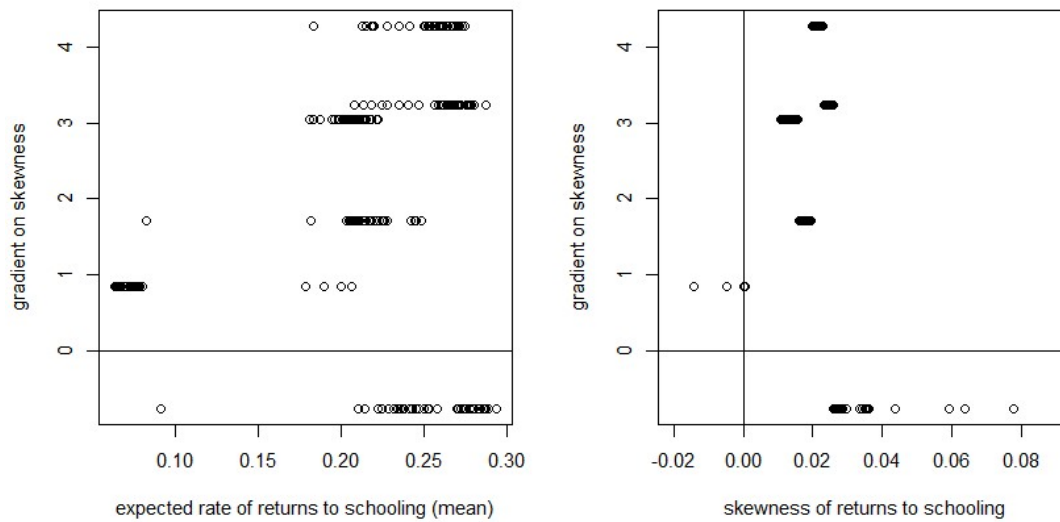**Figure 10.** Spline gradients on Skewness from Step 4.

Figure 11. Spline gradients on skewness from Step 4: Stat. Sign. Only.

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. Educational investment is unique as compared to other types of investments in that rate of returns to education distributions are mostly positively skewed. As we are dealing with mean returns as opposed to log wages, 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 (Hartog and Vijverberg 2007), implying that the more positively skewed a distribution is, the lower its expected returns is going to be on average.
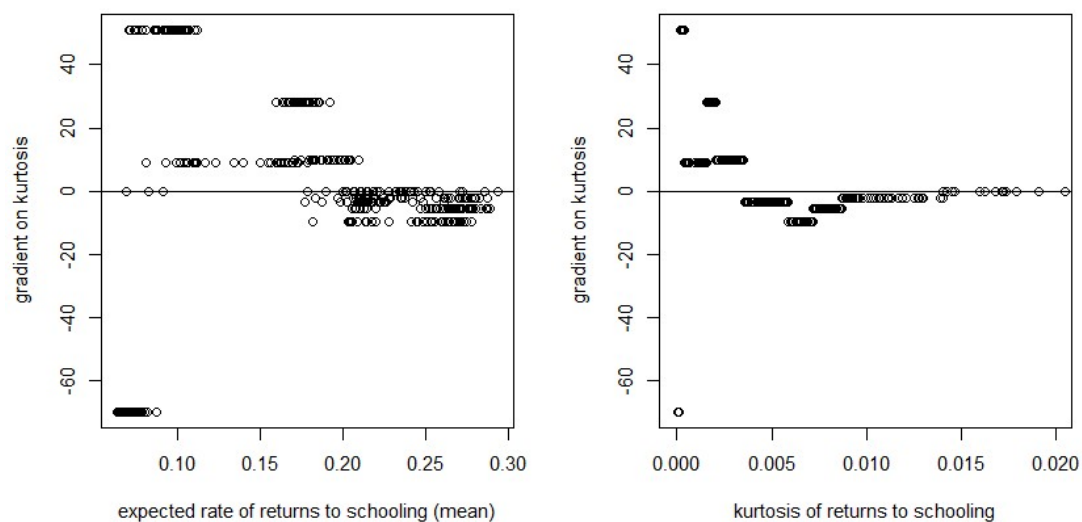

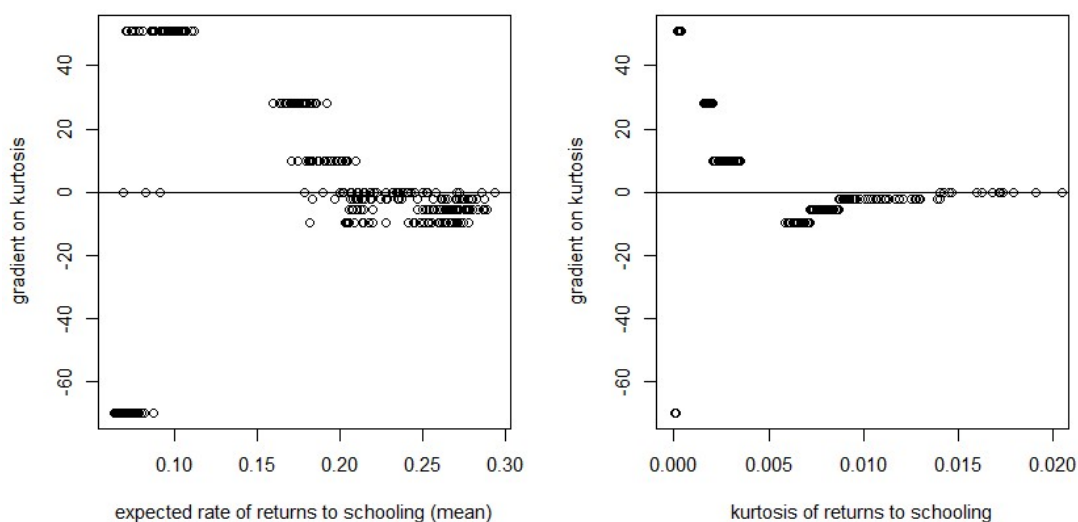**Figure 12.** Spline Gradients on Kurtosis from Step 4.

Figure 13. Spline gradients on kurtosis from Step 4: Stat. Sign. Only.

Our results show that all of the distributions with a negative skewness (upper left quadrant on the right panel of Figures 10 and 11) have positive gradients. This suggests that negative skewness is not compensated by a higher expected rate of return to education as one may expect. If taken literally, this implies that, decreasing skewness from -0.01 to -0.02 (i.e., more risk) would result in a decrease in expected rate of returns to education ($0.85 * -0.01 = -0.0082$). That being said, there are only a few observations with negative skewness (only 5 out of 408 observations are negative), 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 getting a lower return to education on average when they are in a more positively skewed market. The former suggests the opposite, that is, individuals require compensation for positive skewness as well.

The effect of kurtosis on expected returns is nonlinear as well. We see both positive and negative gradients. In the face of log wages we may expect only positive coefficients. While many of these are statistically significant, the kurtosis values themselves are very close to zero. 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 90.81 percentage points. Hence, those percentage changes are large, but the absolute magnitudes are not.

For completeness, Figure 15 plots the gradients on time versus 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).
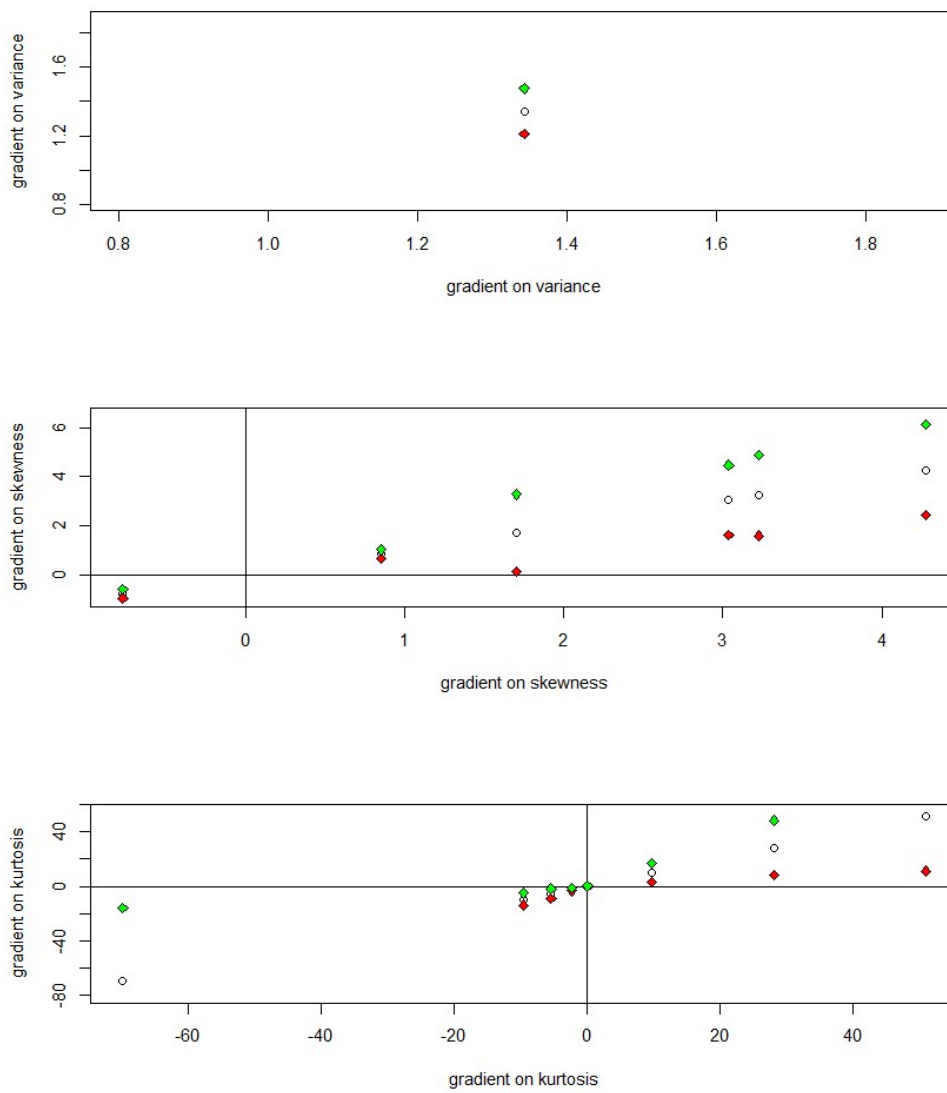
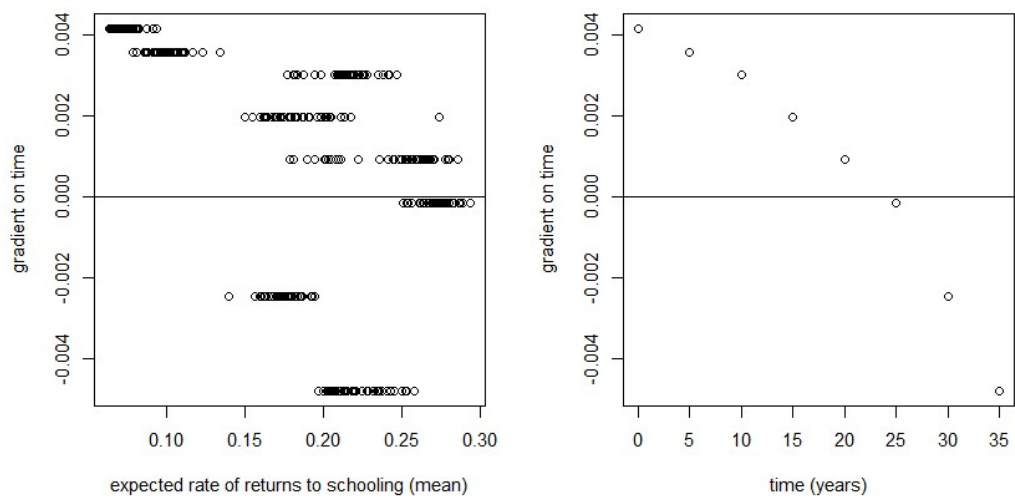**Figure 14.** Statistically significant spline gradients (Step 4) on 45° plot.

Figure 15. Spline gradients on time trend from Step 4.

6. Conclusions

In this paper, we investigated whether education acts as an investment good. If so, parts of individuals' return to education could be a premium for the risk they incur 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 employed nonparametric methods to obtain individual-specific returns to education and hence the entire distribution of returns to education. This approach, therefore, allows us to uncover risk measures beyond variance, including skewness and 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. The compensations for higher-order risks beyond variance are nonlinear with their respective moments, and are not always in the direction (the gradient's sign) that one would expect when the outcome variable is a log wage. While kurtosis has thus far not been studied in an educational investment context, our analysis found that 80% of our gradients on kurtosis are statistically significant. This suggests that kurtosis is also an important risk measure to consider when analyzing educational investment behaviors.

Our results have important implications. First, there is a need to take into account high-order risks in the theoretical modeling of human capital investment, which has been typically ignored in the literature. Second, the literature has typically observed an increasing trend of returns to education. Our results suggest a potential explanation for said pattern is that the trend may partly reflect the increase in various risks surrounding educational investment over time. These results are also important for discussing relevant governmental policies for educational investment, as high estimated returns to education may not necessarily be completely a result of enhanced productivity. This may also explain why individuals choose to drop out of school despite high returns to education. Policies aimed at promoting education should also consider ways to mitigate potential risks associated with educational investment.

A deeper analysis could help better explain the nonlinearities we observed. For example, sub-sampling our final stage further could show possible heterogeneity in risk and in risk compensation. It would be interesting to study the rates of return to education by sex, occupation, race and educational level. It would also be worthwhile to consider the costs of education (e.g., student loans). Future research is also warranted to examine the risk–return trade-off for females, and consider the role of sample selection in estimating the return.

Author Contributions: All authors contributed equally to this manuscript. All authors have read and agreed to the published version of the manuscript.

References

Acemoglu, Daron, and Joshua Angrist. 1999. *How Large Are the Social Returns to Education? Evidence from Compulsory Schooling Laws.* Technical Report. Cambridge: National Bureau of Economic Research.

Angrist, Joshua D., and Alan B. Keueger. 1991. Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106: 979–1014. [CrossRef]

Angrist, Joshua D., and Alan B. Krueger. 1992. 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: 328–36. [CrossRef]

Backes-Gellner, Uschi, Stefan C. Wolter, and Simone N. Tuor. 2010 Risk-return trade-offs to different educational paths: Vocational, academic and mixed. *International Journal of Manpower* 31: 495–519.

Bernoulli, Daniel. 1738. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5: 23–36.

Blackburn, McKinley L., and David Neumark. 1993. Omitted-ability bias and the increase in the return to schooling. *Journal of Labor Economics* 11: 521–44. [CrossRef]

Card, David. 1993. *Using Geographic Variation in College Proximity to Estimate the Return to Schooling*. Technical Report. Cambridge: National Bureau of Economic Research.

Card, David. 1999. The causal effect of education on earnings. In *Handbook of Labor Economics*. Amsterdam: Elsevier, vol. 3, pp. 1801–63.

Card, David. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69: 1127–60. [CrossRef]

Christiansen, Charlotte, Juanna Schröter Joensen, and Helena Skyt Nielsen. 2007. The risk-return trade-off in human capital investment. *Labour Economics* 14: 971–86. [CrossRef]

Cramer, Jan S., Joop A. Hartog, Nicole Jonker, and Mirjam C. van Praag. 2002. Low risk aversion encourages the choice for entrepreneurship: An empirical test of a truism. *Journal of Economic Behavior Organization* 48: 29–36. [CrossRef]

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

Cunha, Flavio, and James J. Heckman. 2007. *The Evolution of Inequality, Heterogeneity and Uncertainty in Labor Earnings in the US Economy*. Technical Report. Cambridge: National Bureau of Economic Research.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault. 2011. Nonparametric instrumental regression. *Econometrica* 79: 1541–65. [CrossRef]

Deschênes, Olivier. 2006. *UNobserved Ability, Comparative Advantage, and the Rising Return to Education in the United States, 1979–2002*. Technical Report, Working Paper, University of California, Santa Barbara, CA, USA.

Eeckhoudt, Louis, and Harris Schlesinger. 2008. Changes in risk and the demand for saving. *Journal of Monetary Economics* 55: 1329–36. [CrossRef]

Ekelund, Jesper, Edvard Johansson, Marjo-Riitta Järvelin, and Dirk Lichtermann. 2005. Self-employment and risk aversion—Evidence from psychological test data. *Labour Economics* 12: 649–59. [CrossRef]

Ghosh, Dipankar, and Manash R. Ray. 1997. Risk, ambiguity, and decision choice: Some additional evidence. *Decision Sciences* 28: 81–104. [CrossRef]

Harmon, Colm, and Ian Walker. 1995. Estimates of the economic return to schooling for the United Kingdom. *American Economic Review* 85: 1278–86.

Harmon, Colm, Vincent Hogan, and Ian Walker. 2003 Dispersion in the economic return to schooling. *Labour Economics* 10: 205–14. [CrossRef]

Harmon, Colm, Hessel Oosterbeek, and Ian Walker. 2003. The returns to education: Microeconomics. *Journal of Economic Surveys* 17: 115–56. [CrossRef]

Hartog, Joop, and Wim PM Vijverberg. 2007. On compensation for risk aversion and skewness affection in wages. *Labour Economics* 14: 938–56. [CrossRef]

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

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2008. Earnings functions and rates of return. *Journal of Human Capital* 2: 1–31. [CrossRef]

Henderson, Daniel J., and Anne-Charlotte Souto. 2018. An introduction to nonparametric Regression for labor economists. *Journal of Labor Research* 39: 355–82. [CrossRef]

Henderson, Daniel J., Solomon W. Polachek, and Le Wang. 2011. Heterogeneity in schooling rates of return. *Economics of Education Review* 30: 1202–14. [CrossRef]

Henderson, Daniel J., Subal C. Kumbhakar, and Christopher F. Parmeter. 2012. A simple method to visualize results in nonlinear regression models. *Economics Letters* 117: 578–81. [CrossRef]

Hoogerheide, Lennart, Joern H. Block, and Roy Thurik. Family background variables as instruments for education in income regressions: A Bayesian analysis. *Economics of Education Review* 31: 515–23. [CrossRef]

Horowitz, Joel L. 2011. Applied nonparametric instrumental variables estimation. *Econometrica* 79: 347–94.

Kaymak, Barış 2009. Ability bias and the rising education premium in the United States: A cohort-based analysis. *Journal of Human Capital* 3: 224–67. [CrossRef]

Kimball, Miles, and Philippe Weil. 1992. *Precautionary Saving and Consumption Smoothing Across Time and Possibilities*. Technical Report. Cambridge: National Bureau of Economic Research.

Kimball, Miles S. 1990. Precautionary Saving in the Small and in the Large. *Econometrica* 58: 53–73. [CrossRef]

Kuznets, Simon, and Milton Friedman. 1939. Incomes from Independent Professional Practice, 1929–1936. In *Incomes from Independent Professional Practice, 1929–1936*. Cambridge: National Bureau of Economic Research.

Ma, Shujie, Jeffrey S. Racine, and Lijian Yang. 2015. Spline regression in the presence of categorical predictors. *Journal of Applied Econometrics* 30: 705–17. [CrossRef]

Mazza, Jacopo, Hans van Ophem, and Joop Hartog. 2013. Unobserved heterogeneity and risk in wage variance: Does more schooling reduce earnings risk? *Labour Economics* 24: 323–38. [CrossRef]

McGoldrick, Kimmarie, and John Robst. 1996. The effect of worker mobility on compensating wages for earnings risk. *Applied Economics* 28: 221–32. [CrossRef]

Meghir, Costas, and Mårten Palme. 2005. Educational reform, ability, and family background. *American Economic Review* 95: 414–24. [CrossRef]

Nie, Zhenghua, and Jeffrey S. Racine. 2012. The crs Package: Nonparametric Regression Splines for Continuous and Categorical Predictors. *R Journal* 4: 48–56 [CrossRef]

Ozabaci, Deniz, and Daniel J. Henderson. 2015. Additive kernel estimates of returns to schooling. *Empirical Economics* 48: 227–51. [CrossRef]

Palacios-Huerta, Ignacio. 2003. An empirical analysis of the risk properties of human capital returns. *American Economic Review* 93: 948–64. [CrossRef]

Parker, Simon C., and C. Mirjam Van Praag. 2006. Schooling, capital constraints, and entrepreneurial performance: The endogenous triangle. *Journal of Business & Economic Statistics* 24: 416–31.

Racine, Jeffrey S., Zhenghua Nie, Brian D. Ripley, and M. J. S. Racine. 2018. Package 'crs'. Avaliable Online: http://kambing.ui.ac.id/cran/web/packages/crs/crs.pdf (accessed on 26 October 2020)

Samuelson, Paul A. 1970. The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *Review of Economic Studies* 37: 537–42. [CrossRef]

Schaafsma, Joseph. 1976. The consumption and investment aspects of the demand for education. *The Journal of Human Resources* 11: 233–42. [CrossRef]

Shaw, Kathryn L. 1996. An empirical analysis of risk aversion and income growth. *Journal of Labor Economics* 14: 626–53. [CrossRef]

Stiglitz, Joseph E. 1974. The demand for education in public and private school systems. *Journal of Public Economics* 3: 349–85. [CrossRef]

Wang, Le. 2013. Estimating returns to education when the IV sample is selective. *Labour Economics* 21: 74–85. [CrossRef]

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