The Effects of Education, Personality, and IQ on Earnings of High-Ability Men

Miriam Gensowski,[1] James Heckman, and Peter Savelyev[2]

The University of Chicago

January 24, 2011

[1]Corresponding author, mgensowski@uchicago.edu.

[2] James Heckman is the Henry Schultz Distinguished Service Professor of Economics at the University of Chicago; University College Dublin; and Yale University. Miriam Gensowski and Peter Savelyev are at the University of Chicago, Department of Economics, as Ph.D. student and Ph.D. candidate, respectively. We are grateful to Min Ju Lee and Molly Schnell for their excellent research assistance. This paper and other versions of it have benefited from discussions at the Labor Working Group, University of Chicago, and the Applied Economics and Econometrics Seminar, University of Mannheim.

Abstract

This paper estimates the internal rate of return (IRR) to education for men and women of the Terman sample, a 70-year long prospective cohort study of high-ability individuals. The Terman data is unique in that it not only provides full working-life earnings histories of the participants, but it also includes detailed profiles of each subject, including IQ and measures of latent personality traits. Having information on latent personality traits is significant as it allows us to measure the importance of personality on educational attainment and lifetime earnings.

Our analysis addresses two problems of the literature on returns to education: First, we establish causality of the treatment effect of education on earnings by implementing generalized matching on a full set of observable individual characteristics and unobserved personality traits. Second, since we observe lifetime earnings data, our estimates of the IRR are direct and do not depend on the assumptions that are usually made in order to justify the interpretation of regression coefficients as rates of return.

For the males, the returns to education beyond high school are sizeable. For example, the IRR for obtaining a bachelor's degree over a high school diploma is 11.1%, and for a doctoral degree over a bachelor's degree it is 6.7%. These results are unique because they highlight the returns to high-ability and high-education individuals, who are not well-represented in regular data sets.

Our results highlight the importance of personality and intelligence on our outcome variables. We find that personality traits similar to the Big Five personality traits are significant factors that help determine educational attainment and lifetime earnings. Even holding the level of education constant, measures of personality traits have significant effects on earnings. Similarly, IQ is rewarded in the labor market, independently of education. Most of the effect of personality and IQ on life-time earnings arise late in life, during the prime working years. Therefore, estimates from samples with shorter durations underestimate the treatment effects.

Contents

1 Introduction

This paper estimates the causal effect of education on earnings, and the corresponding rate of return to education, for a sample of high-ability men. At the same time, we show how measures of personality influence lifetime earnings both directly and indirectly through educational choice.

The Life-Cycle Study of Children with High Ability,[1] hereafter referred to as the "Terman data," is unique and enables us to address several issues which are generally hard to investigate. The study provides rich background information on the participants, as well as detailed measures of personality and interest. Additionally, it is one of the longest prospective cohort studies[2] in existence, so that the subjects' full earnings histories are known. The participants were selected on the basis of having an IQ above 135, and were followed from 1922 to 1991 with surveys every 5–10 years, starting when they were on average ten years old.

We can establish causality of education using the rich set of control variables, including IQ and latent personality traits, and analyze the full lifetime effects on earnings through the longitudinal aspect of the data. For education, the lifetime effect is summarized by the internal rate of return, which we can compute directly instead of just approximating it. For personality, access to the long follow-up means that we can follow the evolution of the role of personality on earnings. Generally, the effects of personality on earnings during the early stages of a career are only a fraction of what they are after age 40.

The Terman sample lets us highlight a few facts: Even in a high-ability group, education has positive returns. At the same time, we do not find much evidence that the production function of marketable skills is convex in ability. Furthermore, we find that even in this top-IQ group, IQ increases earnings directly as well as indirectly through schooling choices. Finally, personality has clear incremental effects on earnings, beyond IQ and schooling.

[1] Terman and Sears (2002a,b); Terman et al. (2002a,b)

[2] Friedman et al. (1995)

We contribute to the literature in three ways. First, we establish causality of education on earnings through a matching procedure. Second, we estimate the *true* internal rate of return to education for the Terman sample, which depends on having the full earnings histories. Third, we show the complete life-time effects of latent personality traits on earnings, which also depends on the availability of a long follow-up.

The first contribution of our paper is the identification of the causal effect of education, which we establish with a matching procedure. Matching is made possible by the richness of the Terman data. Establishing causality between education and higher earnings has been at the center of many efforts in the field. If unobservable variables (most prominently, "ability") positively influence the amount of education obtained and are independently rewarded, then the coefficient on education overstates the direct causal effect of education on earnings when one fails to control for such variables. Since schooling is the result of optimizing behavior, and since the decision maker has more information about himself than the economist, the amount of schooling that an individual obtains is likely to be endogenous.[3] Matching assumes that the researcher has all relevant control variables at his disposal, and that the participant's

[3]The two main approaches to deal with the issue of endogeneity of the schooling decision discussed in Card (1999) are instrumental variables and twin studies. The IV analyses are based on the premise that there are variables that affect the cost of schooling, but not the benefit. They exploit institutional sources of variation in schooling, i.e. the minimum school leaving age, tuition costs, or the geographic proximity of colleges. The most well-known articles are by Angrist and Krueger (1991), Staiger and Stock (1997) (discussed by Bound et al. (1995) and Bound and Jaeger (1996)), Kane and Rouse (1993), Card (1995), and Angrist and Krueger (1992). The instruments used vary greatly in quality – see Carneiro and Heckman (2002) for a critical analysis. IV estimates are generally larger than the corresponding OLS estimates. It is unclear how to interpret the point estimates, because with heterogeneous returns to schooling, the point estimate of the schooling coefficient is a weighted average of the idiosyncratic marginal benefits for the persons whose schooling choices were affected by the instrument. Papers using twins or brothers are based on the fact that twins or brothers share at least one common component of the unobservable ability variable. For the most prominent works after Chamberlain and Griliches (1975) and the survey by Griliches (1979), see the papers on monozygotic twins by Ashenfelter and Krueger (1994); Ashenfelter and Rouse (1998); Rouse (1999), and Behrman and Taubman (1994). Measurement-error corrected within-family estimates were slightly smaller than the simple OLS coefficients. The remaining question is whether twins' abilities and skills are truly identical, especially considering personality traits. These factors are important inputs into schooling and valued in the labor market (Bowles et al., 2001; Heckman et al., 2006). Therefore, estimates of the effect of schooling on wages that control for genetic background factors can still be biased. Furthermore, research by Fraga et al. (2005) and others shows that even in monozygotic twins, epigenetic differences arise during the lifetime. Heckman and Vytlacil (2007) explicitly discuss structural models (the control function approach), matching, and other treatment effect estimators, and how they can be put into a common framework of marginal treatment effects.

selection into treatment can be represented as a selection on these variables.

The Terman data is optimal for estimating the treatment effect with matching. As we argue below, the sample itself is quite homogenous. In addition, the data provides a large set of essential observable background variables: the respondent's childhood health, parental background, family environment, and teenage health. The data also includes the respondent's cognitive ability (IQ) and measures of latent personality traits. Controlling for IQ addresses the ability bias concern. "Psychic costs" are considered to be another important determinant of schooling choice, and economists generally cannot measure these costs. We approximate psychic costs by explicitly taking into account measures of personality, such as conscientiousness and extraversion. These traits are highly relevant for both schooling choice and earnings, and are thus an integral part of the matching procedure.[4]

The second contribution we make is to compute the internal rate of return instead of only interpreting the coefficient on schooling in a wage equation as "the rate of return." Even though the return to education is one of the central ideas in labor economics,[5] most articles on the subject do not actually estimate the internal rate of return. The coefficient on years of schooling in a Mincer equation, which is hedonic in nature, can only be interpreted as the

[4] The Terman data has been used extensively by psychologists, but only scarcely by economists. Known to us are only Becker et al. (1977); Hamermesh (1984); Michael (1976); Tomes (1981) and Leibowitz (1974). The only economic paper that uses the Terman sample to analyze the effect of education attainment on earnings is Leibowitz (1974). This paper estimates a Ben-Porath model of investment in human capital, in the pre-school, school, and post-school periods. Income is modeled as being derived from the rents on human capital in the form of ability and home, schooling, and postschooling investments. The estimation, however, deals with ability and home investments separately, not in one equation. It is in fact an earnings equation at three points in time (1940, 1950 and 1960). "Home investments" are only proxied by parents' education and family income. The measure of schooling is "years of schooling," converted from the categorical data on degrees obtained. Including childhood IQ as a covariate in the wage equation does not alter the coefficient on education by much. The coefficient on years of schooling ranges from .063 (in 1940) to .075 (in 1960). Estimates are very similar for the three OLS specifications - standard Mincer regression with only schooling and a quadratic in experience; controlling for parents' education and family income; or controlling for childhood IQ. Our paper has a different goal, and uses a more refined set of control variables than Leibowitz. We do not claim to estimate a Ben-Porath investment model, and we use earnings from all years instead of only a few.

[5] It was rendered popular by Mincer (1974) who estimated Becker and Chiswick (1966)'s model, in which the coefficient on schooling could indeed be interpreted as a rate of return. Belzil and Hansen (2002) note "A World Wide Web survey of the most recent literature indicates that, since 1970, more than 200 published articles or working papers (set in a reduced-form) have been devoted to the estimation of the return to schooling or surrounding issues." For reviews of these estimations in the literature, see Psacharopoulos (1981), Psacharopoulos and Patrinos (2004), Willis (1986), and Card (1999).

rate of return under stringent conditions: linearity of earnings in years of schooling, constant working life, and no explicit or psychic costs of college. These assumptions have been tested and rejected by Heckman et al. (2006, 2008). In our analysis, we compute the internal rate of return to obtaining one degree versus another instead of assuming that earnings are linear in years of schooling. The earnings histories in the Terman data are complete, so we know the length of the working life. The explicit cost of college is taken into account,[6] as well as tax rates by marital status.[7] The indirect effects of education on earnings, for example through marriage, longevity and the labor-leisure choice, are also accounted for in our analysis. See Figure 1 for the average earning of males and females by level of education.

Our third contribution is to show how the effect of personality on earnings varies throughout the men's working lives. We find that without access to long follow-up data, the estimated effect would be understated. Note that even though the Terman sample has a restricted range of IQ, there is substantial variation in personality. In fact, the Terman men do not differ from the general population in terms of personality.

The paper proceeds as follows. First, we describe the matching approach, and argue how this identifies the causal effect of education on earnings. In Section 2 we describe the difference between the internal rate of return as computed in our approach from what is often used in the literature. Then, in Section 3, we present the estimates of the rates of return to all education pairs. Section 4 addresses the effect personality has on life-time earnings, both directly and indirectly through education.

[6]From a detailed education history that includes the name of the college or university attended, we impute the cost of schooling. The participants also gave information on scholarships and fellowships, which is taken into account in this computation. Note that the costs considered here are purely pecuniary and exclude psychic cost, for example.

[7]The tax rates and corresponding brackets are taken from form US-1040 by the IRS, collected by the Tax Foundation at http://www.taxfoundation.org/publications/show/151.html. We use the marital status at each age, as determined by the marriage history we construct, in order to apply different tax rates for singles and married participants. Unfortunately, we only have one measure of income, so the tax brackets are determined based on earnings (or family earnings) only. This possibly understates the participants' tax dues, if they had substantial non-wage income.

2 The Model

Our empirical analysis relies on matching to identify the causal effect of education on earnings. The invoked matching assumptions guarantee identification of the average treatment effect. For a discussion of the method, the potential outcomes representation is useful. Model each person's outcomes (i.e. his earnings) in two states 0 and 1 as

$$
\begin{aligned}
Y_1 &= \mu_1(X,\theta) + \varepsilon_1, \\
Y_0 &= \mu_0(X,\theta) + \varepsilon_0.
\end{aligned}
$$

Here, treatment (state 1) denotes the higher education level. Therefore, Y_1 corresponds to earnings one would have with higher education, and Y_0 corresponds to earnings one would have with lower education. $\mu_k(X,\theta)$ is the expected mean earnings in treatment state k, conditional on observed background variables X and latent variables θ. Note that Y_1 and Y_0 are *potential* outcomes only; they cannot be both observed for the same person. Let D indicate the treatment. Then, we observe

$$
\begin{aligned}
Y &= DY_1 + (1-D)Y_0 \\
&= Y_0 + D(Y_1 - Y_0) \\
&= \boldsymbol{\mu_0}(X,\boldsymbol{\theta}) + \varepsilon_0 + D(\boldsymbol{\mu_1}(X,\boldsymbol{\theta}) + \varepsilon_1 - \boldsymbol{\mu_0}(X,\boldsymbol{\theta}) - \varepsilon_0).
\end{aligned}
$$

When a person has the higher schooling level, $D = 1$ and we observe Y_1. When he has the lower schooling level, $D = 0$ and we observe Y_0. In the potential outcome approach, a treatment's impact is given by the comparison of the observed outcome to the other, counterfactual, outcome: $\Delta = Y_1 - Y_0$. However, we only observe outcome $Y = DY_1 + (1-D)Y_0$, and thus there exists an evaluation problem (we observe one individual in only one of the possible treatment states). Also, there is a selection problem since individuals select into treatment based on potential outcomes. Therefore,

$$(Y_0, Y_1) \not\perp\!\!\!\perp D$$

and $E(Y_1 \mid D = 1) - E(Y_0 \mid D = 0) \neq E(Y_1 - Y_0)$.

2.1 Matching Assumptions

Matching assumes that conditioning on observables X eliminates the dependence between (Y_0, Y_1) and D. The two matching assumptions are

$$(Y_0, Y_1) \perp\!\!\!\perp D|X \tag{M-1}$$

$$0 < \Pr(D = 1|X) < 1. \tag{M-2}$$

Propensity score matching, a well-known variant of matching, reduces the dimensionality. The propensity score $P(X) = \Pr(D = 1|X)$ is the probability of participation, or the probability of obtaining the higher education level. Rosenbaum and Rubin (1983) prove that when the two matching conditions (M-1) and (M-2) hold, we can also express the first as

$$(Y_0, Y_1) \perp\!\!\!\perp D|P(X). \tag{M-1'}$$

Now assume that we can observe, or have access to, otherwise latent variables θ. Using these variables in the set of conditioning variables allows us to relax assumption (M-1). With such information, the following is more appropriate:

$$(Y_0, Y_1) \perp\!\!\!\perp D|X.\theta \tag{M-1”}$$

(M-1”) can again be modified to condition on $P(X, \theta)$ instead of X and θ separately.

Based on the potential outcome model outlined above, the treatment effect at each age is

$$\Delta_t = Y_{1,t} - Y_{0,t} = \boldsymbol{\mu_1}(X_t, \boldsymbol{\theta}) - \boldsymbol{\mu_0}(X_t, \boldsymbol{\theta}) + \varepsilon_{1,t} - \varepsilon_{0,t}. \tag{1}$$

The choice of modeling $\boldsymbol{\mu_1}(X_t, \boldsymbol{\theta})$ and $\boldsymbol{\mu_0}(X_t, \boldsymbol{\theta})$ remains. The functional form we will employ for this paper is the common coefficient model,[8] where outcomes are modeled as

$$\begin{aligned}\mu_1 (X_t, \theta) &= c_{1,t} + X_t\beta_t + \theta\delta_t,\\ \mu_0 (X_t, \theta) &= c_{0,t} + X_t\beta_t + \theta\delta_t.\end{aligned}$$

Estimation of the treatment effect consists only of regressing the observed Y on the observed X, latent θ, and the treatment indicator:

$$Y_t = X_t\beta_t + \theta\delta_t + D\overline{\Delta}_t + c_{0,t} + e.$$

Here, $\overline{\Delta}_t$ is the average treatment effect of D on Y_t at time t. In the case of multiple treatment states, notably the five education levels from high school diploma to doctoral degree, D is a matrix of treatment indicators.

Our measures of θ are the predicted factor scores. Thus, we introduce error by using these factor scores rather than the true factors. However, knowing the factor model, we can characterize the measurement error and correct the coefficients on the factors accordingly.

Matching does not model the decision process. It relies on the data being sufficient to

[8]This choice is the result of a tradeoff between flexibility and measurement error correction. Nonparametric matching does not make the strong functional form assumptions as linear separable parametric forms do. The latter versions, on the other hand, allow for a correction of attenuation bias. Measurement error is introduced into our estimation by using predicted factor scores instead of the true factors. However, the precise form of the attenuation bias is known, and thus can be corrected using the covariance matrix of the true factors that are computed during the factor estimation. Heckman et al. (2010) (Web Appendix G) describe this correction method. In Appendix B, we present treatment estimations that test three different functional forms for $\boldsymbol{\mu}(X_t, \boldsymbol{\theta})$: local linear matching (kernel matching), a linear separable model for each treatment state, and the common coefficient model. Interestingly, the respective treatment effect estimates are almost identical. Therefore, we chose the common coefficient model which is computationally very simple and allows for the measurement error correction.

make the decision variable conditionally independent from the distribution of outcomes.[9] Assumption (M-2) can be verified, and is usually no cause of debate (we verify it in Appendix B.1). Assumption (M-1) (statistical conditional independence), however, implies that conditional on X, the marginal return equals the average return. This is a strong behavioral assumption, and as Heckman and Vytlacil (2007) note, "Many economists do not have enough faith in their data to invoke it." The Terman data, however, allows us to match subjects much more closely than can usually be done. We now argue why.

2.2 IQ and Personality Factors

We match not only on observable variables, but also on latent traits. Most models concerned with ability bias in education are of the "single-factor" type, with an underlying hierarchical interpretation of ability. We account for multiple types of ability, notably IQ and a vector of personality traits. Denote the vector of these traits θ.

IQ was measured at study entry in 1922, and was the basis for inclusion in the Terman sample. Most students took the Stanford-Binet IQ score. About 30% of the students took another IQ test, the "Terman Group Test" (for a more detailed description of the tests, see Chapter I in Terman and Sears, 2002a). The data gives us only one IQ score, so in order to control for possible differences in measurement in our analysis, we include an indicator for the "Terman Group Test" and an interaction of the score with this indicator.

We define the included latent personality traits similarly to the Big Five taxonomy, notably Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). While our traits are conceptually very close to the Big Five personality traits, they are not measured with the same inventory. However, inspection of the items shows the close correspondence. Furthermore, Martin and Friedman (2000) have shown that Conscientiousness and Extraversion from the Terman questionnaires correspond closely to the Big Five traits.

[9] Structural models of the schooling decisions have been estimated by, for example, Keane and Wolpin (1997), Eckstein and Wolpin (1999), or Belzil and Hansen (2002). The technology of skill formation has been modeled in Cunha and Heckman (2007); Cunha et al. (2006). Hansen et al. (2004) show how schooling and ability measures interact.

To quantify the personality traits, we compute factor scores using a three-step estimation procedure, as outlined in Heckman et al. (2010). This estimation procedure, which will be explained in more detail below, extracts the factors from personality ratings in 1922, 1940, and 1950. We use teacher-, parent-, and self-ratings from these different surveys in order to cover more aspects of the men's personality. Some items are only available in the later years, while others are only given in the survey at the beginning of the study.[10]

The two personality traits from 1922 are Openness and Extraversion. The factor score for Extraversion is extracted from ratings of the subject's "fondness for large groups," "leadership," and "popularity with other children." The factor score for Openness is extracted from ratings of the subject's "desire to know," "originality," and "intelligence." [11] The personality ratings are averages of both parents' and teachers' ratings of the subject's display of the indicated personality traits. All ratings are given on a scale from 1 to 13, ranging from extraordinarily low/ poor to extraordinarily high/ good (i.e. higher numbers are associated with better traits).

The dedicated items for the factors of Conscientiousness, Agreeableness, and Neuroticism are based on self-ratings in 1940 and 1950. Where items from both years are available, we use the average. In a few cases, we use mean-imputation for missing item responses. The self-ratings of personality traits from 1940 and 1950 are on an 11-point scale, with 11 being the high end of the trait described. In 1940, the subjects filled out an extensive list of

[10] See our discussion of whether personality can be considered a causal factor in the determination of wages. This caveat clearly applies more so for the factor scores constructed from the self-ratings in 1940 and 1950 than for the teacher and parent ratings of 1922, since these are pre-market ratings. The fact that these ratings are so early constitutes at the same time a drawback: contemporaneous personality determines wages, not the personality of when individuals were 10 years old. Personality may evolve over the life cycle, and by using early measures one disregards this possibility.

[11] Due to the phrasing, it might seem as if Openness and IQ measure the same underlying trait. Note that the IQ test is a direct test of the subject's cognitive ability, while the parents' and teachers' ratings describe their impressions of the child. Furthermore, several measurements of these impressions combine to the factor defined by psychologists as "Openness." Hogan and Hogan (2007) define the Big Five Openness as the degree to which a person needs intellectual stimulation, change, and variety. Openness is indeed correlated with IQ, at .16 (significantly different from zero), but not perfectly. IQ is negatively related to Extraversion (-.12), also statistically significant. The correlations between IQ and the other personality traits are not statistically significant. The only significant correlations among the personality factors are between Extraversion and Openness (.27), Extraversion and Agreeableness (.09), Neuroticism and Conscientiousness (-.11), and Neuroticism and Extraversion (-.09).

personality items of the Bernreuter personality inventory. These items are questions about usual behavior and feelings that can be answered "yes," "no," and "?."

Conscientiousness is constructed from self-ratings of "persistence" and "definite purposes," as well as the Bernreuter items "Do you enjoy planning your work in detail?" and "In your work do you usually drive yourself steadily?" Neuroticism is based on the measurements on "moodiness," "sensitive feelings," "feelings of inferiority," and the Bernreuter items "Are you much affected by the praise or blame of many people?," "Are you frequently burdened by a sense of remorse or regret?," "Do you worry too long over humiliating experiences?," "Are you feelings easily hurt?" Agreeableness is based on the ratings of "easy to get along with" as well as the Bernreuter items "Do you usually try to avoid arguments?," "Are you always careful to avoid saying anything that might hurt anyone's feelings?," "Do you often ignore the feelings of others when doing something that is important to you?".[12]

In this paper, we use the three-step estimation procedure outlined in Heckman, Malofeeva, Pinto, and Savelyev (2010). In the first step, we use the measures for the multiple components of each personality factor to estimate the parameters of each factor's respective measurement system. We denote measures that capture factor j by $M^j_{m^j}$, where $m^j \in \mathcal{M}^j$. There may be a different number of measures for each of the factor types $j \in \mathcal{J}$. The measurement systems are of the form

$$\begin{aligned} M^j_1 &= \nu^j_1 + \theta^j + \eta^j_1, \\ M^j_{m^j} &= \nu^j_{m^j} + \varphi^j_{m^j}\theta^j + \eta^j_{m^j}; \qquad m^j \in \mathcal{M}^j \setminus \{1\}, \forall j \in \mathcal{J}, \end{aligned}$$

where the factor loading associated with the first measure of each factor is normalized to unity to set the scale of the factors. At the end of the first step, these parameters are used to predict factor scores by the Bartlett (1937) method. In the second step, we use the factor scores as covariates in our matching analysis. Finally, in the third step, we adjust the

[12]For more information about the factors and selection of measures, see Savelyev (2010).

coefficients on the factor scores for the bias introduced by using the predicted factor score instead of the true factor. The intuition behind this adjustment is similar to the standard attenuation-bias formula. The formula for the attenuation through the predicted factor is a function of the covariance matrix of the true factor as well as the measured factor score. The latter is readily available, and the former can be extracted from the factor estimation itself.

These personality traits are both predictors of educational choice and explanatory variables in the wage outcome equations. Their effects on education and wages will be discussed in detail in Section 4.

2.3 Matching Variables

In addition to the availability of high quality measures of latent factors, there are two other elements of the Terman data which render it ideal for matching: a very homogenous sample and the availability of a large set of relevant observable background variables.

Subjects are already approximate matches due to the homogeneity of the sample. All subjects are highly intelligent and were living in California at the time of the study's inception. They are Caucasian and generally lived in advantageous environments (the vast majority are from middle-class families).

Second, the Terman data provides a large number of covariates, allowing us to control for a wide array of essential variables that influence both education and labor market success. Respondents are matched on IQ score at the beginning of the study,[13] father's and mother's backgrounds (education, occupation, social status, region of origin, age at birth of subject), family environment (family's finances when growing up, number of siblings, birth order), and early childhood health (birthweight, breastfeeding, sleep quality in 1922). Additional controls are birth cohort group (birth year 1904-10 or 1911-15), and whether the subject was

[13] For most subjects, this was the Stanford Binet IQ Score. Some of the participants, however, took the "Terman Group Test," a test that was specifically designed for screening these high achieving children. We control for potential differences between the effects of these tests by including a dummy that indicates the Terman Group Test, and an interaction between the recorded IQ score and this indicator.

active in combat in World War II.[14] Table 1 presents descriptive statistics for the sample and all background variables used.

Third, we control for the fundamental latent personality types of the individuals, as described in section 2.2. Personality traits are not only highly relevant to the educational choice, but also influence earnings directly.

2.4 The Internal Rate of Return

Equipped with the treatment effect at each age, the internal rate of return (IRR) is easily computed. Age t ranges from 18 to 75. The IRR is the discount rate that equates lifetime earnings streams for two different schooling levels (a non-marginal difference). The IRR, ρ, is defined as the solution to the following polynomial:

$$\sum_{t=18}^{75} \frac{\overline{\Delta}_t}{(1+\rho)^{t-17}} = 0 \tag{2}$$

where $\overline{\Delta_t} = E[Y_{1,t} - Y_{0,t}|X_t] = ATE$.[15]

Instead of finding ρ, the net present value (NPV) assumes a fixed interest rate, r, and reports the discounted sum of earnings differences. For the purposes of our examples, we use a discount rate of 5% whenever we report the NPV.

While the IRR is useful as a summary of an investment project (in a single number), it has certain shortcomings. When the IRR is compared to the current interest rate, one can "read off" whether the investment should be undertaken – as long as one is only interested in the profitability of *one* project, and as long as this project's cash flow changes from positive to negative only once. For comparison of two mutually exclusive projects, the NPV is a better guide. It does not suffer from the scale problem and the timing problem as the IRR

[14] While there are more covariates available, in order to avoid overfitting, we selected a group of the most relevant characteristics and background variables.

[15] Note that we are interested in all direct and indirect effects that schooling has on *lifetime earnings*. These comprise effects of education on the labor-leisure choice (including retirement or unemployment) and longevity. Since we are not estimating a pricing equation of human capital, we use a comprehensive earnings measure (annual earnings in levels) that also reflects the intensity with which human capital is used.

does (for a discussion and examples, see Chapter 6 of Ross, Westerfield, and Jaffe (2001)). Finally, note that neither the NPV nor the IRR take into account uncertainty, psychic costs of college, or differential costs of funding.

interpreted as the product of early childhood investments and early childhood human capital. Then, as schooling may influence IQ, later IQ-advances would be picked up by the schooling coefficient. The coefficient on schooling would r

3 The Returns to Education

We will now discuss the returns to education for the men in the Terman sample. This section first takes as given the personality traits, which are used as covariates in the matching procedure, but not discussed explicitly. Section 4 then presents how personality influences life-time earnings. Personality traits affect total earnings directly since they are rewarded in the market place. Furthermore, personality affects educational choices, and thus through the returns to these educational achievements personality also indirectly influences earnings.

3.1 Treatment Effects of Education, Pairwise IRRs

Recall that the treatment effect of education tells us, in a counterfactual sense, how much a person would have gained or lost as a result of obtaining more or less education. The treatment effects of all education pairs are shown in Figures 2 to 6. The effect describes by how much having the higher degree, in comparison to the lower level of education, improves average earnings at each age, *holding everything else constant*. The treatment effect of higher education is negative in the men's early years, since those obtaining higher education are still attending school while their peers with less education are already out of school and in the job market. Later, during the prime working years, the positive effect of education is substantial. This is a standard result in the literature. The outcome variable is annual earnings after tax and tuition, in 2008 U.S. Dollars. The tax rates used are a function of marital status (married or single). Tuition was subtracted from earnings at each year that college was attended, at both the undergraduate and graduate level.[16]

The IRRs and NPVs corresponding to the treatment effects are summarized in Table 3. In comparison to having a high school diploma, obtaining a bachelor's degree increases earnings by $111,788 over a lifetime, if the difference in earnings is discounted at 5%. The corresponding IRR is 11.1%. In comparison to a high school diploma, having completed only some college courses leads to only slightly higher earnings throughout one's life, but since the investment costs are very low, the corresponding rate of return of 9.0% seems relatively high. Since the investment period for obtaining a master's degree or a doctoral degree is longer than for a bachelor's degree, the rates of return for these education levels in comparison to a high school diploma are lower than the 11.1% figure from the bachelor's degree. The IRRs are 8.0% for a master's degree and 8.9% for a doctoral degree over a high school diploma. Note that at almost identical rates of return, the doctoral degree nevertheless leads to much higher discounted earnings gains than the master's degree ($79,867 vs $144,491). The rates of return of having a college degree or higher in comparison to "some college" are almost equivalent to the returns over "high school only." The difference between the two base-line education levels only appears in present value terms — the discounted gains in comparison to "some college" are around $25,000 lower than in comparison to "high school diploma." Note that this is similar to the earnings difference between high school and some college. In comparison to having a bachelor's degree, having a master's degree has almost no return. However, obtaining a doctoral degree over a bachelor's degree does increase lifetime earnings, corresponding to an IRR of 6.7% and a present value of the difference of $32,703. This NPV seems rather low because most of the gains arise late in the working life and are discounted heavily. The return to having a doctorate over a master's degree is high (12.5%). In this

[16] For tuition rates, we drew on de Gruyter, W., ed. (1948) and Hurt, H., ed. (1949) from 1920 to 1940. Details on how the tuition data was constructed are given in Appendix A.3. We have made two implicit assumptions about tuition payments: 1. by subtracting them at the time of college attendance, we exclude smoothing out of the expense; and 2. we assume that graduate students paid the full tuition as noted in the aforementioned sources. Our results change in only minor ways when we relax both assumptions.

case, both groups have relatively long investment periods, but men with doctoral degrees have higher earnings. Thus, in this comparison, the investment is low and the return is high.

As explained in the previous section, the IRRs should not be used for determining the optimality of one education investment versus another, in comparison to a third education level which is the baseline.[17] In principle, one compares one IRR to the prevalent interest rate. However, this type of comparison ignores the dynamic aspect of schooling and the sequential revelation of uncertainty.[18] Our analysis is explicitly ex-post and considers rates of return in a static setting.

We can draw two conclusions from these results. One is that even in a very high ability group, education adds skills that are valued in the marketplace. The returns to schooling are real, and ability bias cannot be responsible for the type of returns we find. The second conclusion is that there is little evidence for a convexity of the production function of skills in ability. If there was such a convexity (that is, more able individuals learn more from school than less able individuals), we would have expected higher returns than those we observe.

[17]For the reader that is startled by the "nonlinear" pattern of some of the pairwise IRRs, let us consider an example. For example, if for males the return of a master's degree versus a bachelor's degree is only 1.2%, how can it be that the IRR of getting a doctoral degree versus a master's degree is 12.5%, but a doctoral degree versus a bachelor's degree is only 6.7%? Shouldn't the two IRRs be more similar, and if anything the IRR of a doctorate versus a master's degree a little lower? To understand these numbers, examine the graphs of the pairwise treatment effects. We see that initially, as they pursue more schooling, those with a master's degree have negative treatment effects. These negative effects are only barely offset by slightly higher earnings late in life. Thus, even though there is a difference between the two earnings streams, the IRR is very small. Now if we compare a doctoral degree to a bachelor's degree, the pattern is similar, except that the men with a doctoral degree have a sizeable positive treatment effect later in life. Thus, the IRR is greater than for a master's degree versus a bachelor's degree. But if we proceed to the comparison of a doctorate versus a master's degree, note that men in both groups will spend more time in school, and thus forego earnings. In comparison to the men with a master's degree, the men with a doctorate are *not* losing out as much as in comparison to one with a bachelor's degree who start earning earlier. However, those with doctoral degrees will proceed to have substantially higher earnings. Thus, since there are almost no initial costs of getting a doctorate in comparison to getting a master's degree, but sizeable gains, the IRR of a doctorate versus a master's degree is sizeable (12.5%).

[18] See for example Heckman et al. (2006) for a discussion of the problems and particularities associated with sequential resolution of uncertainty. The option value of schooling has been analyzed, for example, by Heckman and Urzua (2008).

3.2 Comparison to Estimates from Simpler Methods

How much would our estimates of the rates of return change if we did not have access to the personality factors and IQ?

Part a) of Table 4 shows results from the matching procedure without these covariates. We still include the full set of background variables, but we do not include the latent personality traits or IQ. The left half of Table 4 shows the NPVs from this specification, and the right half shows the bias in the NPVs from this reduced regression. They are overstated in all cases, and the bias can get as large as 98%.

Note that there is evidence of omitted variable bias when we exclude personality measures and IQ. Omitting personality and IQ is akin to, but slightly different from, ability bias in the traditional sense. The difference is that in the relatively IQ-homogenous Terman sample, omtting IQ does not bias the IRRs substantially. Separate analyses (not shown) prove that omitting the personality factor scores leads to a greater bias than omitting IQ. The modest bias in terms of IRRs that we find from omitting both personality measures and IQ is related to the way in which IRRs are estimated. The treatment effects in the prime-working years (age 40–60) are actually decreased by including the IQ and personality factors, but, due to discounting, the IRR does not pick up much of these later changes. The NPVs, on the other hand, do reflect the higher treatment effects. Note furthermore, in a preview of results in Section 4.3, that while the treatment effects are not greatly affected by the inclusion of personality factors and IQ, these variables do significantly influence wages directly.

One might have access to a similar longitudinal cohort study as the Terman data, where one is reasonably confident that individuals come from similar backgrounds. However, the availability of such a rich set of background characteristics to control for is very rare. What would happen if one attempted to use the matching procedure outlined in this paper but with fewer matching variables? To simulate such a case, we apply the matching procedure on the Terman data using only state of birth, two cohort indicators, and parental education as covariates. As the results in part b) of Table 4 show, all of the NPVs from this estimation are biased upwards as well. Clearly, the small subset of regressors misses sources of variation in both education and earnings. The gaps between the treatment effect curves are particularly large for the doctoral degree versus bachelor's degree and master's degree versus bachelor's degree comparisons.

4 The Effects of IQ and Personality Traits on Lifetime Earnings

So far, we have focused on the returns to education, using personality and IQ only as control variables. However, personality and IQ are clearly interesting in their own right, and this section deals with these variables explicitly. We ask very generally "How do personality and IQ affect life-time earnings?" After briefly analyzing the overall effect of personality on total earnings, we focus on the two main channels: 1. personality traits affects one's educational attainment and thus affect wages indirectly through education, and 2. personality traits are rewarded independently in the labor market and thus affect wages directly. Section 4.2 presents results on the role of personality traits in educational choice, and Section 4.3 shows the "gains to personality" holding education constant.

4.1 The Total Effect of Personality and IQ on Lifetime Earnings

We begin by analyzing how personality and IQ influence lifetime earnings. We use the sum of each individual's earnings from age 18 to age 75.[19] The first column of Table 5, "Total Effect," exhibits coefficients from the regression of lifetime earnings on personality traits and IQ only. The coefficients reflect a very general association between the personality variables in the Terman data and the male's lifetime earnings. No other covariates were controlled for. With this simple regression, Conscientiousness and Extraversion are positively associated

[19]Here, we use the undiscounted sum of earnings, but a separate analysis with discounted earnings (at, for example 5%,) shows that all results presented here are maintained.

with earnings, while Agreeableness and Openness are negatively associated with earnings (although Openness fails to be statistically significant in this very simple exercise). Our measure of Neuroticism does not have a clear association with earnings. It is remarkable that even in this very high-IQ sample, where the range of observed IQs is clearly restricted, IQ still has a positive and statistically highly significant association with lifetime earnings.

We call these simple associations "total effect" of latent personality traits and IQ on lifetime earnings since these traits affect lifetime earnings both indirectly through educational attainment (which we will explore more in the next section) and directly. We have already shown in Section 3.1 that schooling influences earnings substantially, independently of personality. Therefore, not conditioning on schooling in the regression of lifetime earnings on personality traits subsumes the effect of schooling in the personality variables' coefficients.

The second column, "Total Effect, with covariates" adds the full set of control variables. We thus control for background characteristics which might be correlated with personality, as well as schooling (still kept implicit). However, the estimates remain very similar.

Finally, the third column, "Direct Effect, given Education" presents the effect of personality on lifetime earnings, holding education constant. Again, it includes all covariates from the treatment effect analysis (parametric matching described in Section 2). The base line is "Doctoral Degree," and in comparison to this education level, all other educational categories have clearly lower lifetime earnings. The role of the personality traits and IQ are preserved. Conscientiousness and Extraversion still have large and positive effects on life-time earnings. Agreeableness has a negative lifetime "return", conditional on education.

Finally, note that even when controlling for rich background variables, IQ maintains a statistically significant effect on lifetime earnings. Even though the effect is slightly diminished from the un-controlled association of the first column, it is still sizeable. Malcolm Gladwell claims rather generally in his book "Outliers" that for the Terman men, IQ did *not* matter once family background and other observable personal characteristics were taken into account. While we do not want to argue that IQ has a larger role for the difference between

50 and 100, for example, than for the difference between 150 and 200, we do want to point out that even at the high end of the ability distribution, IQ has meaningful consequences.

One caveat about causality is in order. In contrast to the causal effect of education on earnings, there is a risk of reverse causality in the analysis of the effect of personality on earnings. Most researchers use early measures of personality and analyze the effects of these early measures on later outcomes, thus being certain that there is no reverse causality. We partially follow this approach by using early measures of Openness and Extraversion. However, the other personality traits are measured at a time where the men are already in their working lives. Thus, these measures are more relevant to the observed earnings, but at the same time we cannot exclude the possibility that, for example, a high score on Neuroticism is a *result* of one's position in the workforce. Therefore, while we do think of the results as showing earnings gains *due to* personality and IQ, we do not claim causality as we do in the case of education.

Another point of discussion concerns the role of personality in economic models. In this paper, we have followed the standard practice of just including personality variables as covariates in regressions, without modelling the manifest personality explicitly. Instead of this "standard methodology" one could model observed traits as a response to utility maximization under constraints, such as suggested in Almlund et al. (2011).

4.2 The Effects of IQ and Personality on Education

Several authors have analyzed how personality traits, and notably the Big Five factors, influence years of schooling obtained. For example, Almlund et al. (2011) summarize existing evidence on this relationship in datasets that are representative of the entire population (for the U.S., the Netherlands, and Germany). In these populations, Conscientiousness is always positively associated with years of education, while Extraversion, Agreeableness, and Neuroticism are negatively associated. While Openness exhibits a positive association, this effect is probably due to the correlation between Openness and IQ, which is not controlled for in these samples. When we run a simple regression of *years of schooling* on the personality variables and IQ, as well as the full set of background variables (not shown), we find that Conscientiousness unambiguously increases years of schooling. We can add to the results from the aforementioned representative samples that even at the high end of intelligence, Conscientiousness is still a separate and statistically significant predictor of schooling attainment. With IQ and the other additional background controls, the other personality factor scores are not statistically different from zero.

We believe, however, that discrete educational choices and degrees, the way we have analyzed them in the section on returns to schooling, are more meaningful than years of schooling completed. This holds all the more so at the high end of the educational spectrum. Therefore, we analyze educational choice using a multinomial logit model, which allows for five separate categorical outcomes. Having a high school diploma is the base level. Table 6 presents the relative risk ratios associated with each educational outcome in comparison to the baseline.

IQ has a positive influence on schooling attainment. A higher IQ in this sample makes it more likely to obtain higher schooling than a high school diploma. This is interesting given the highly selected sample we have; even for individuals with an IQ of 140 and over, having a higher IQ makes it less likely to remain in the lowest schooling category. Yet, this does not necessarily imply that having a higher IQ increases the odds of obtaining a master's degree rather than a bachelor's degree.

Conscientiousness predicts obtaining more education than a high school diploma as well; while only the comparison with the doctoral degree is statistically significant. It is intuitive to interpret Conscientiousness as lowering the psychic costs of education. In addition, the "future planning" element of Conscientiousness can be thought of as lowering the discount rate of future gains (recall that these gains are substantial, but accrue late in life). Also, a greater tendency to plan for the future could decrease the effort needed to imagine future outcomes and to correctly evaluate the costs and gains involved in the long-term investments of obtaining higher education. The effect of Conscientiousness is not only highly significant, but also linear. This means that higher Conscientiousness is always positive for educational attainment. This stands in contrast to the effects of the following two personality traits.

Neuroticism is only associated with one outcome, namely with *not* being in the category of "some college." Men who score higher on the Neuroticism scale are much less likely to be in this category than in the base outcome. Since this factor score is not significant for any other schooling level comparisons, we interpret this finding to mean that only men who are relatively stable emotionally remain in the vague schooling category without a college degree. The lack of importance of Neuroticism in the determination of schooling is somewhat surprising, especially in light of evidence from other sources. Notably, Locus of Control is often linked to Neuroticism, and an internal Locus of Control has been found by many authors to reliably predict higher schooling (see, for example Piatek and Pinger, 2010 or Baron and Cobb-Clark, 2010).

Extraversion seems negatively related with education, but in a non-linear fashion. Having a higher extraversion score decreases the odds of obtaining more than a high school diploma, everything else held constant, but only to a certain degree.

The factor scores on Openness and Agreeableness do not produce relative risk ratios that are statistically significantly different from 1 in the multinomial logit with high school as the base level. However, we can see that having a low score on Openness and a high score on Agreeableness would tend to increase the probability of either remaining in the "some college" category, or obtaining a doctoral degree.

We know that education has positive returns. Therefore, through the choice of educational level, personality indirectly affects lifetime earnings.

4.3 The Effects of IQ and Personality on Wages

How are earnings affected by personality traits, given the educational level? We have seen already in Table 5 that earnings and measures of personality traits are highly correlated.

Since we are also interested in investigating when the "gains to personality" arise in the working life, a more detailed analysis is useful. In the treatment effect computations of Section 3, we controlled for the latent personality traits and IQ. Here, we discuss the coefficients on the factor scores that were obtained in these regressions. Specifically, we are interpreting the coefficients δ in equation (2.1). We interpret the coefficients as the direct effect of personality on wages. The effect of personality traits on educational attainment is controlled for by directly including the education indicators.

As expected, IQ has a positive and statistically significant effect on earnings for much of the life cycle (Panel a) of Figure 8. The effect of IQ, even in this select sample, is never negative. The positive gains from IQ start accruing relatively early in the working life, at age 30.

Openness has a negative effect on earnings in the Terman sample. Until age 40, there is no effect of Openness on earnings at all, only late in the working life does the negative effect materialize. Note that individually, the effects are not statistically significantly different from zero, although the direction is clearly negative.

Conscientiousness and Extraversion have the largest effects on earnings. Both traits clearly have a "return" in the sense that more conscientious and more extroverted individuals have higher earnings in the labor market, holding the level of education constant. As with IQ, the returns to these character traits are positive throughout the life cycle, although the largest gains appear in the prime working years, ages 45-55. Thus, if researchers have access only to earnings observations for the early working life, the gains from these two personality traits would likely be understated. Also note that while Conscientiousness increases earnings *directly and indirectly*, Extraversion has two different effects on lifetime earnings: It decreases them indirectly through its association with lower educational attainment, but increases earnings directly. Since the total effect (Table 5) of Extraversion on lifetime earnings is positive, the indirect effect must be small in comparison to the direct effect.

Agreeableness has a negative effect on earnings. As with Openness, these negative effects are mostly in the later working years. Had we analyzed the effect of personality on earnings only up to age 35, we would have completely missed this negative effect.

In our sample, Neuroticism appears to have no effect on earnings. This is in line with findings from other datasets. For example, Piatek and Pinger (2010) show using the German SOEP that Locus of Control, does not influence wages when they control for education.

Note that the effects we just discussed are restricted to be linear, since the factor scores enter the treatment effect analysis only in levels. This restriction might be masking underlying non-linear patterns. In fact, one would almost expect the effects of personality on earnings to be non-linear. For example, being somewhat extroverted might be beneficial for one's career, but being *extremely* extroverted might influence inter-personal relations in such a way that a career is actually hindered by this degree of Extraversion.

However, when we analyze the quadratic terms of the personality factor scores, we find that there is no pervasive evidence for strong nonlinearities. The quadratic terms are not significantly different from zero. Figures 11 to 13 give a graphical representation of the potential nonlinear effects by quartiles of each personality factor score.

As shown in Figure 11, Conscientiousness has a positive marginal effect on earnings at all quartiles. Similarly, the marginal effect of Extraversion is positive even at the 75th percentile. Nevertheless, there is some evidence for a slightly decreasing marginal "return" to Extraversion, as the marginal effect at the first quartile is generally greater than at the third quartile.[20] As shown in Figure 12, the effect of Openness on earnings, by quartile, looks very similar to the linear effect of Openness as well. Agreeableness displays rather interesting nonlinear effects. In the earlier years of the working life, up to age 35, the linear effect was close to zero. It seems as if in this range, there are negative marginal "returns". The marginal effect at the highest levels of Agreeableness is now *negative* while at the

[20]The careful reader might notice that the linear effect of Figure 11 is a little larger than we would expect from the three quartiles depicted in Panel a) of Figure 9. The coefficients for the linear effect are all corrected for the attenuation bias from the measurement error introduced by using predicted factor scores. Unfortunately, this correction is not possible in the quadratic term (since we cannot determine the 4th moment of the true factors). Therefore, the estimates for the quadratic term are still attenuated and are absolutely smaller than the corrected coefficients.

lowest levels, it is (slightly) positive. Beyond age 40, however, this relationship reverses: the marginal effect of becoming more agreeable at already high levels of agreeableness is slightly positive (or zero), while the effect from becoming slightly more agreeable from very low levels is *negative*. The marginal effect of Agreeableness on earnings is convex. Neuroticism (Figure 13) displays a concave effect in the negative range: increasing the Neuroticism score from very low levels to higher levels (corresponding to going from being emotionally very stable to being somewhat less stable) decreases earnings greatly. Once a certain level of Neuroticism is reached, however, increasing it further has almost no effect on earnings. Finally, IQ also displays decreasing marginal "returns," although the effect remains clearly positive at all quartiles of the IQ distribution.

As a note, tests of interaction between the effect of personality traits and IQ and education have not indicated any heterogeneity. The interaction terms are never statistically significantly different from zero. The effect of personality traits, in the Terman sample, seems to be working through levels only.

Clearly, personality both drives educational choice and helps explains wage differences within a given education level.

5 Conclusion

This paper estimates the true internal rate of return to education and the present discounted value of education for the high-achieving men of the Terman study without relying on the strong assumptions that are typical in the literature. We establish causality through matching on an unusually extensive list of covariates. Observing the full earnings history ex post and having access to background variables of this high quality is unique. Returns at the high end of education can be analyzed well with this sample, and we find that such returns are still sizeable.

Personality traits are also shown to have significant and meaningful effects on earnings.

Conscientiousness and Extraversion have positive effects on earnings both directly and indirectly through increasing educational attainment. Other traits, such as Agreeableness, have positive indirect effects but negative direct effects. The direct reward to these traits materializes mostly in the prime working years, not during the early part of one's career.

References

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wössmann (Eds.), *Handbook of the Economics of Education*. Amsterdam: Elsevier.

American Council on Education, ed. (1967–1979). A Fact Book on Higher Education.

Angrist, J. D. and A. B. Krueger (1991, November). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics 106*(4), 979–1014.

Angrist, J. D. and A. B. Krueger (1992, May). Estimating the payoff to schooling using the vietnam-era draft lottery. NBER Working Papers 4067, National Bureau of Economic Research, Inc.

Ashenfelter, O. and A. B. Krueger (1994, December). Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review 84*(5), 1157–1173.

Ashenfelter, O. and C. Rouse (1998, February). Income, schooling, and ability: Evidence from a new sample of identical twins. *Quarterly Journal of Economics 113*(1), 253–284.

Baron, J. D. and D. Cobb-Clark (2010). Are Young Peoples Educational Outcomes Linked to their Sense of Control? Discussion Paper 4907, Institute for the Study of Labor (IZA), Bonn.

Bartlett, M. S. (1937, July). The statistical conception of mental factors. *British Journal of Psychology 28*(1), 97–104.

Becker, G. S. and B. R. Chiswick (1966, March). Education and the distribution of earnings. *American Economic Review 56*(1/2), 358–369.

Becker, G. S., E. M. Landes, and R. T. Michael (1977). An economic analysis of marital instability. *Journal of Political Economy 85*(6), 1141–1187.

Behrman, Jere R., M. R. R. and P. Taubman (1994). Endowments and the allocation of schooling in the family and the marriage market: the twins experiment. *Journal of Political Economy 102*, 1131–1174.

Belzil, C. and J. Hansen (2002, September). Unobserved ability and the return to schooling. *Econometrica 70*(5), 2075–2091.

Bound, J. and D. A. Jaeger (1996, November). On the validity of season of birth as an instrument in wage equations: A comment on angrist and krueger's "does compulsory school attendance affect schooling and earnings?". NBER Working paper 5835, National Bureau of Economic Research, Inc.

Bound, J., D. A. Jaeger, and R. M. Baker (1995, June). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Bowles, S., H. Gintis, and M. Osborne (2001, December). The determinants of earnings: A behavioral approach. *Journal of Economic Literature 39*(4), 1137–1176.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, and R. Swidinsky (Eds.), *Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp*, pp. 201–222. Toronto: University of Toronto Press.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 5, pp. 1801–1863. New York: North-Holland.

Carneiro, P. and J. J. Heckman (2002, October). The evidence on credit constraints in post-secondary schooling. *Economic Journal 112*(482), 705–734.

Chamberlain, G. and Z. Griliches (1975, June). Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers. *International Economic Review 16*(2), 422–449.

Conrad, H. (1956). Trends in tuition charges and fees. Higher Education.

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V. Masterov (2006). Interpreting the evidence on life cycle skill formation. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 12, pp. 697–812. Amsterdam: North-Holland.

de Gruyter, W., ed. (1928, 1936, 1940, 1948). American Universities and Colleges.

Eckstein, Z. and K. I. Wolpin (1999, November). Why youths drop out of high school: The impact of preferences, opportunities, and abilities. *Econometrica 67*(6), 1295–1339.

Fraga, M. F., E. Ballestar, M. F. Paz, S. Ropero, F. Setien, M. L. Ballestar, D. Heine-Suer, J. C. Cigudosa, M. Urioste, J. Benitez, M. Boix-Chornet, A. Sanchez-Aguilera, C. Ling, E. Carlsson, P. Poulsen, A. Vaag, Z. Stephan, T. D. Spector, Y.-Z. Wu, C. Plass, and M. Esteller (2005, July). Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America 102*(30), 10604–10609.

Friedman, H. S., J. S. Tucker, J. E. Schwartz, L. R. Martin, C. Tomlinson-Keasey, D. L. Wingard, and M. H. Criqui (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology 68*(4), 696–703.

Griliches, Z. (1979, October). Sibling models and data in economics: Beginnings of a survey. *Journal of Political Economy 87*(5), S37–S64.

Hamermesh, D. S. (1984). Life-cycle effects on consumption and retirement. *Journal of Labor Economics 2*(3), 353–370.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004, July–August). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1-2), 39–98.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997, October). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies 64*(4), 605–654.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings equations and rates of return: The Mincer equation and beyond. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 7, pp. 307–458. Amsterdam: Elsevier.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2008, Spring). Earnings functions and rates of return. *Journal of Human Capital 2*(1), 1–31.

Heckman, J. J., L. Malofeeva, R. Pinto, and P. A. Savelyev (2010). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Heckman, J. J. and S. Urzua (2008, June). The option value of educational choices and the rate of return to educational choices. Unpublished manuscript, University of Chicago. Presented at the Cowles Foundation Structural Conference, Yale University.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4875–5144. Amsterdam: Elsevier.

Hogan, R. and J. Hogan (2007). *Hogan Personality Inventory Manual* (3rd ed.). Tulsa, OK: Hogan Assessment Systems.

Hollingworth, L. S., L. M. Terman, and M. Oden (1940). *NSSE Yearbook, Intelligence: Its Nature and Nurture. Comparative and Critical Exposition*, Volume 39, Chapter The significance of deviates, pp. 43–89. National Society for the Study of Education.

Hurt, H., ed. (1923, 1928, 1933, 1939, 1949). The College Blue Book.

Kane, T. J. and C. E. Rouse (1993, January). Labor-market returns to two- and four-year colleges: is a credit a credit and do degrees matter? NBER Working paper 4268, National Bureau of Economic Research, Inc.

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Leibowitz, A. (1974, March/April). Home investments in children. *Journal of Political Economy 82*(2), S111–S131.

Martin, L. R. and H. S. Friedman (2000). Comparing personality scales across time: An illustrative study of validity and consistency in life-span archival data. *Journal of Personality 68*(1), 85–110.

Martin, L. R., H. S. Friedman, and J. E. Schwartz (2007). Personality and mortality risk across the life span: The importance of conscientiousness as a biopsychosocial attribute. *Health Psychology 26*(4), 428–436.

Michael, R. T. (1976). Factors affecting divorce: a study of the terman sample. NBER Working Papers 147, National Bureau of Economic Research, Inc.

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: Columbia University Press for National Bureau of Economic Research.

Piatek, R. and P. Pinger (2010). Maintaining (locus of) control? Assessing the impact of locus of control on education decisions and wages. Discussion Paper 5289, Institute for the Study of Labor (IZA), Bonn.

Psacharopoulos, G. (1981, October). Returns to education: An updated international comparison. *Comparative Education 17*(3), 321–341.

Psacharopoulos, G. and H. A. Patrinos (2004, August). Returns to investment in education: A further update. *Education Economics 12*(2), 111–134.

Rosenbaum, P. R. and D. B. Rubin (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Ross, S. A., R. W. Westerfield, and J. Jaffe (2001, October). *Corporate Finance* (6th ed.). McGraw-Hill/Irwin.

Rouse, C. E. (1999). Further estimates of the economic return to schooling from a new sample of twins. *Economics of Education Review 18*, 149–157.

Savelyev, P. (2010). Personality, education, and longevity of high-ability individuals. Unpublished manuscript, University of Chicago, Department of Economics.

Snyder, T. D. (Ed.) (1993). *120 Years of American Education: A Statistical Portrait.* Center for Education Statistics, U.S. Department of Education.

Staiger, D. and J. H. Stock (1997, May). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Terman, L. M. (1939, October). Educational suggestions from follow-up studies of intellectually gifted children. *Journal of Educational Sociology 13*(2), 82–89.

Terman, L. M. and R. R. Sears (2002a). *The Terman Life-Cycle Study of Children with High Ability, 1922-1986*, Volume 1, 1922-1928. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Terman, L. M. and R. R. Sears (2002b). *The Terman Life-Cycle Study of Children with High Ability, 1922-1986*, Volume 3, 1950-1986. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Terman, L. M., R. R. Sears, L. J. Cronbach, and P. S. Sears (2002a). *The Terman Life-Cycle Study of Children with High Ability, 1922-1986*, Volume 2, 1936-1945. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Terman, L. M., R. R. Sears, L. J. Cronbach, and P. S. Sears (2002b). *The Terman Life-Cycle Study of Children with High Ability, 1922-1986*, Volume 4, 1991, Parts 69 and 70. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Tomes, N. (1981, April). A model of fertility and children's schooling. *Economic Inquiry 19*(2), 209.

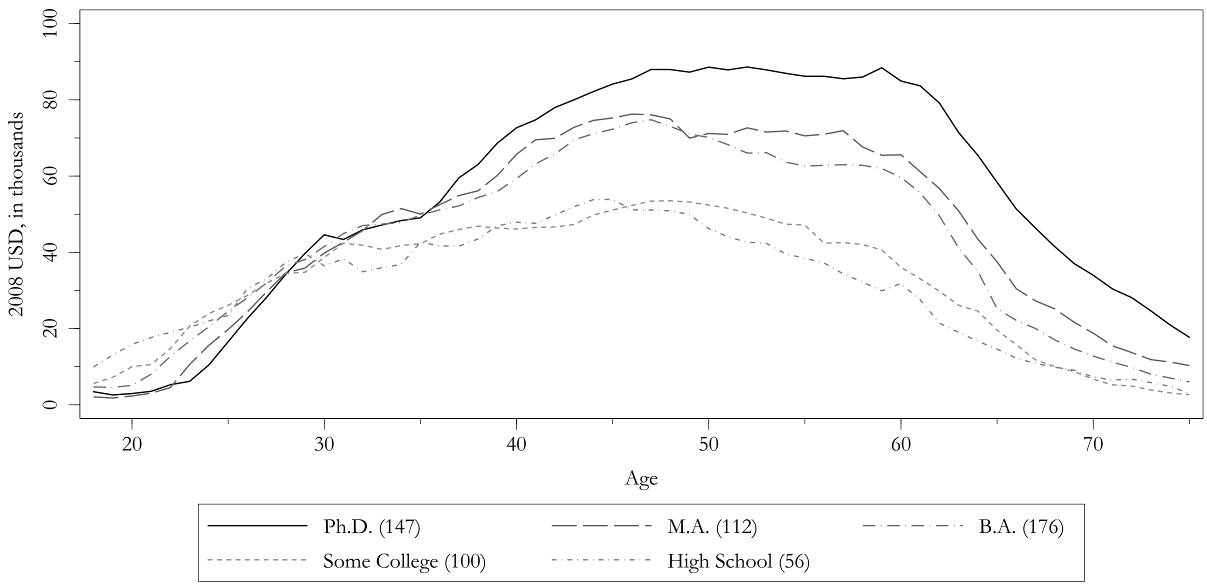
von Lösch, A.-M. (1999). *Der nackte Geist: die Juristische Fakultät der Berliner Universität im Umbruch von 1933*. Mohr Siebeck.

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1, pp. 525–602. New York: North-Holland.

6 Tables and Figures

(Page intentionally left blank.)

Figure 1: Average Earnings by Education, Minus Tuition and After Taxes



Notes: Observation counts are given in parentheses. Earnings are average annual earnings after tax and minus tuition, in 2008 U.S. Dollars, constructed from Terman Data. The tax rates and brackets used are for singles and married persons according to marital status. The tuition cost is applied in full when it occurred, i.e. we do not assume any smoothing out of the payment streams, and we assume graduate students pay full tuition as well. The sample is the same as for the treatment effect computation. The education categories refer to the highest educational level attained in life. See Appendix A for information on building the earnings profiles and tuition from the raw data.

Table 1: Descriptive Statistics of the Terman Sample used, Part I

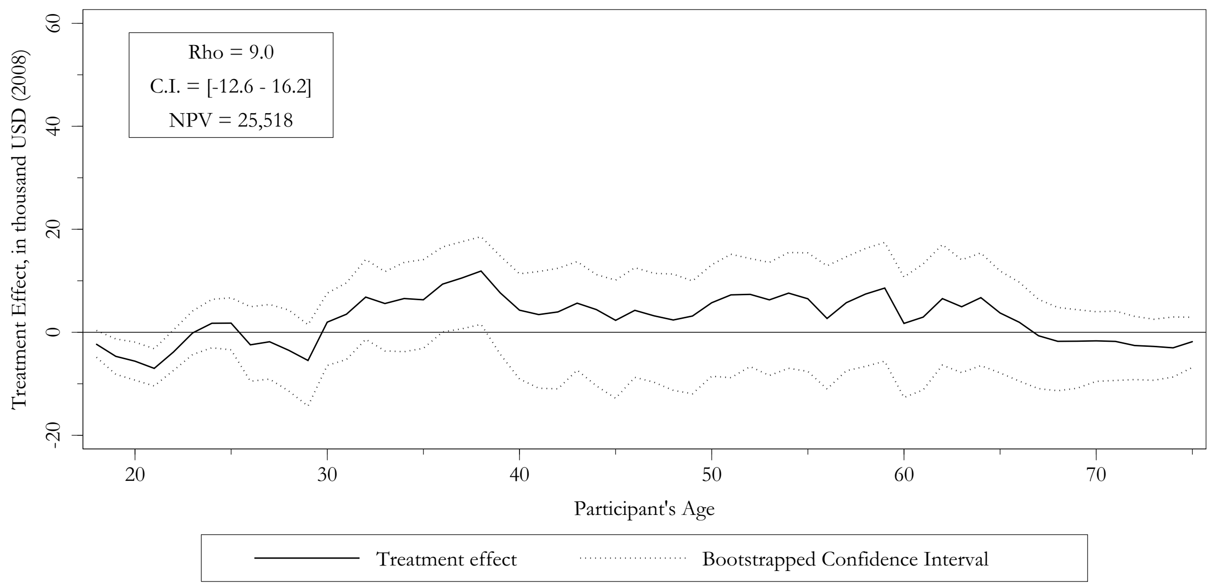
Variable		**Males**	
Education Levels	**Year**	**#Obs**	**%**
High school	1991	59	9.6
Some college	1991	102	16.5
Bachelor's/ some graduate	1991	182	29.5
Master's or equivalent	1991	115	18.6
Ph.D. or equivalent	1991	159	25.8
Basic Information	**Year**	**Mean**	**Std.Dev**
Conscientiousness	1940, 1950	3.51	.83
Openness	1922	3.49	.83
Extraversion	1922	3.51	.68
Agreeableness	1940, 1950	3.50	.60
Neuroticism	1940, 1950	3.50	.61
IQ normalized	1922	5.02	1.01
Terman Group Test	1922	1.38	2.06
Outcomes			
Married at age 30	1934 - 1946	.74	.44
Married at age 40	1944 - 1956	.87	.33
Length of life in 1993	1993	43.41	38.01
Parental Background			
Father's occupation: clerical or deceased	1922	.27	.44
Father's occupation: low-skilled	1922	.17	.37
At least one parent is retired or deceased	1922	.03	.18
Mother has occupation (not minor)	1922	.12	.32
Father's age when child was born	1922	.57	.50
Mother's age when child was born	1922	28.71	5.40
Mother's age at birth, <25		.21	.41
Mother's age at birth, >=35		.14	.35
Father's highest school grade	1922	10.66	5.53
Father's HSG: at most 9 yrs		.39	.49
Father's HSG: 10-13 years		.26	.44
Father's HSG: at least 14 yrs		.34	.48
Mother's highest school grade	1922	10.32	4.50
Mother's HSG: at most 4 yrs		.11	.32
Mother's HSG: 5-11 years		.38	.49
Mother's HSG: at least 12 yrs		.51	.50

Table 2: Descriptive Statistics of the Terman Sample used, Part II

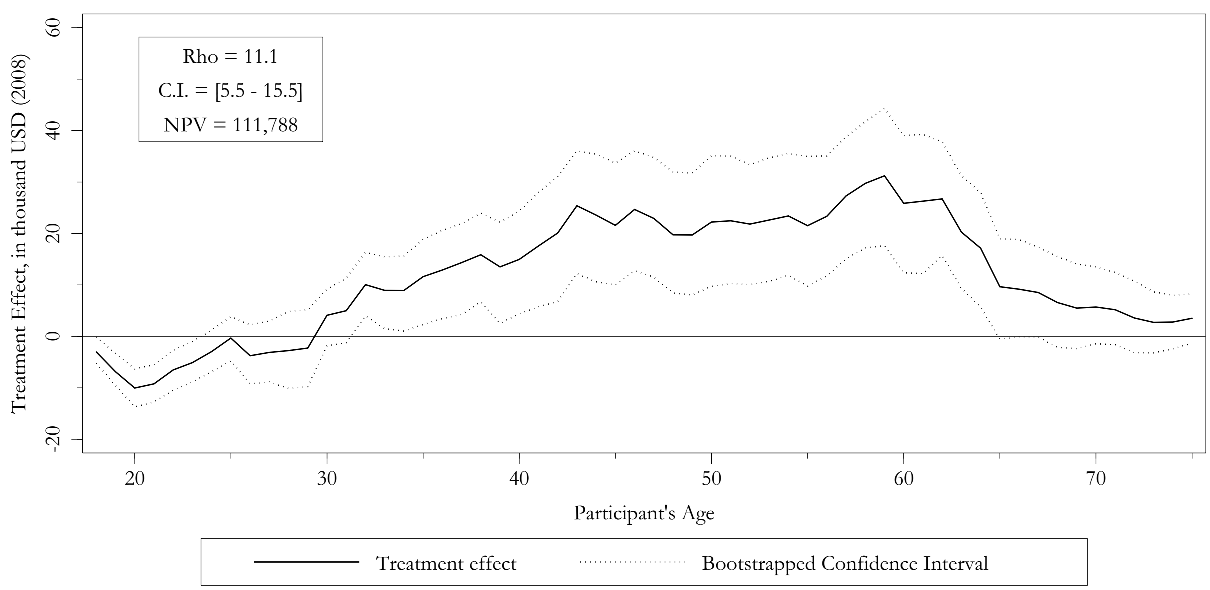
Variable		**Males**	
Parental Background (continued)	**Year**	**Mean**	**Std.Dev**
Either parent is born in Europe	1922	.13	.34
Childhood family finances - (very) limited	1950	.37	.48
Childhood family finances - abundant	1950	.04	.20
Childhood parental social status - high	1950	.34	.47
Siblings			
Number of siblings	1940	1.59	1.61
No sibling		.15	.36
2-4 siblings		.37	.48
5-9 siblings		.06	.24
Birth order	1940	1.84	1.26
Birth order: 2		.22	.41
Birth order: 3		.11	.32
Birth order: 4 +		.20	.40
Early Health			
No breastfeeding	1922	.10	.29
Birthweight in kilograms	1922	3.80	.66
Sleep is sound	1922	.97	.17
Cohort Information			
Cohort: 1904-1910		.24	.43
Cohort: 1911-1915		.47	.50
WWII combat experience	1945	.10	.29

Figure 2: Pairwise Treatment Effects on After-Tax Earnings, Males

(a) Some College vs High School



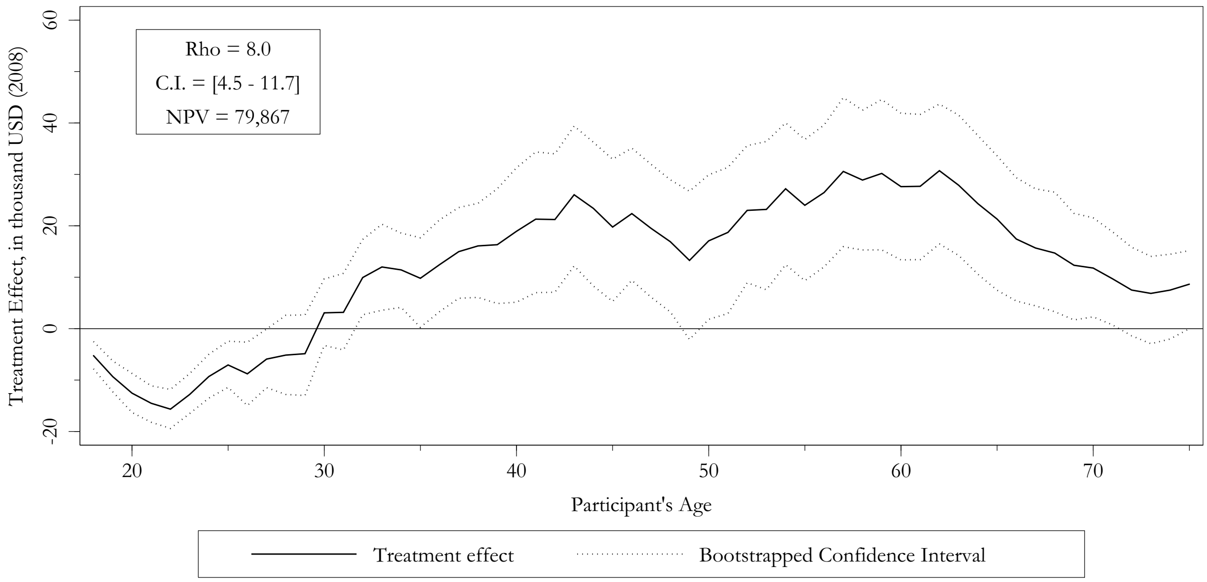
(b) Bachelor's vs High School



Notes: Treatment effects from the common coefficient model. Dotted lines are the 90% basic bootstrap confidence interval, based on 200 bootstrap draws. Earnings are average annual earnings after tax, in 2008 U.S. Dollars. The tax rates and brackets used are for singles and married persons according to marital status. The tuition cost is applied in full when it occurred, i.e. we do not assume any smoothing out of the payment streams, and we assume graduate tudents pay full tuition as well. The covariates are IQ, factor scores for Conscientiousness, Neuroticism, Agreeableness, Openness, and Extraversion, parental background, family environment, early childhood health and 1922 health information, and controls for WWII and cohort. See Notes to figure B-2 or text for more details. See Appendix A for information on building the earnings profiles, and the marriage history, from the raw data.

Figure 3: Pairwise Treatment Effects on After-Tax Earnings, Males

(a) Master's vs High School

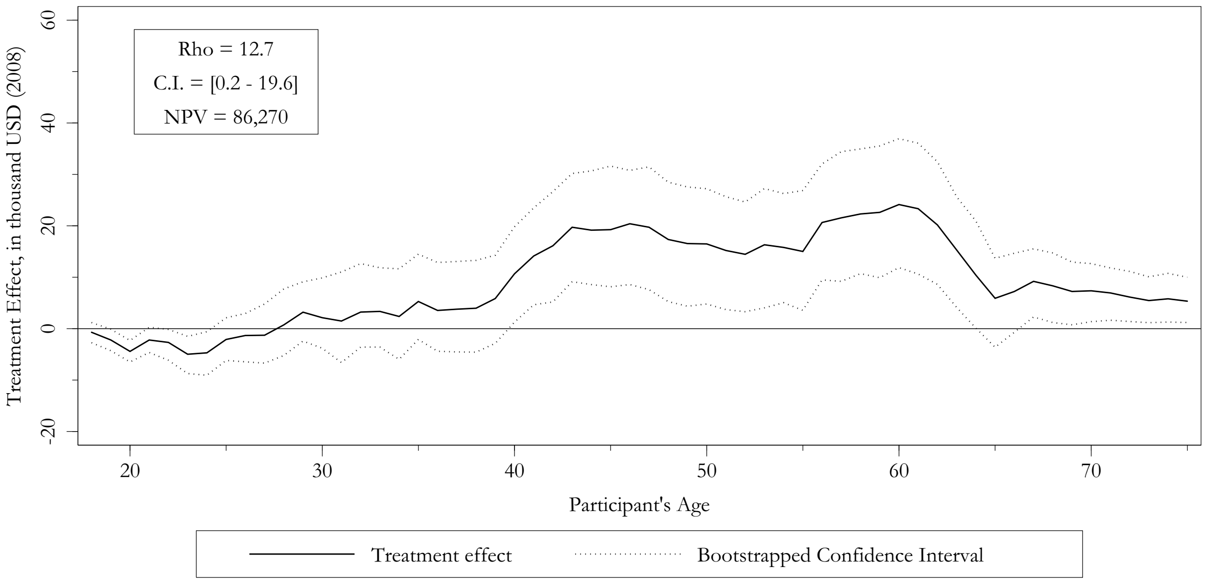


(b) Doctorate vs High School

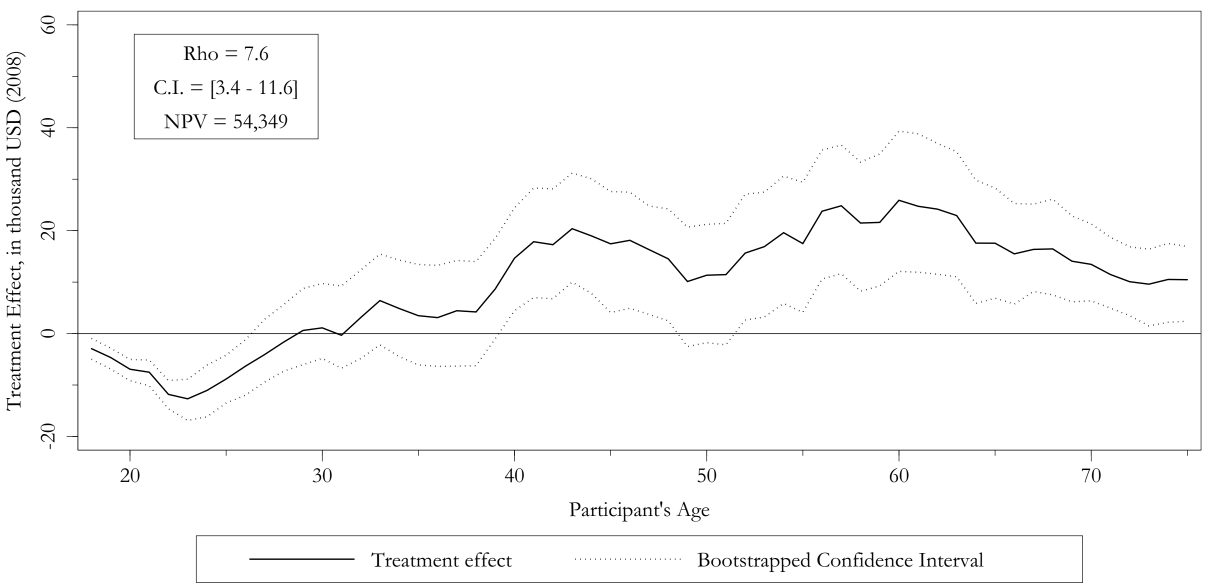
Notes: See notes to Figure 2.

Figure 4: Pairwise Treatment Effects on After-Tax Earnings, Males

(a) Bachelor's vs Some College



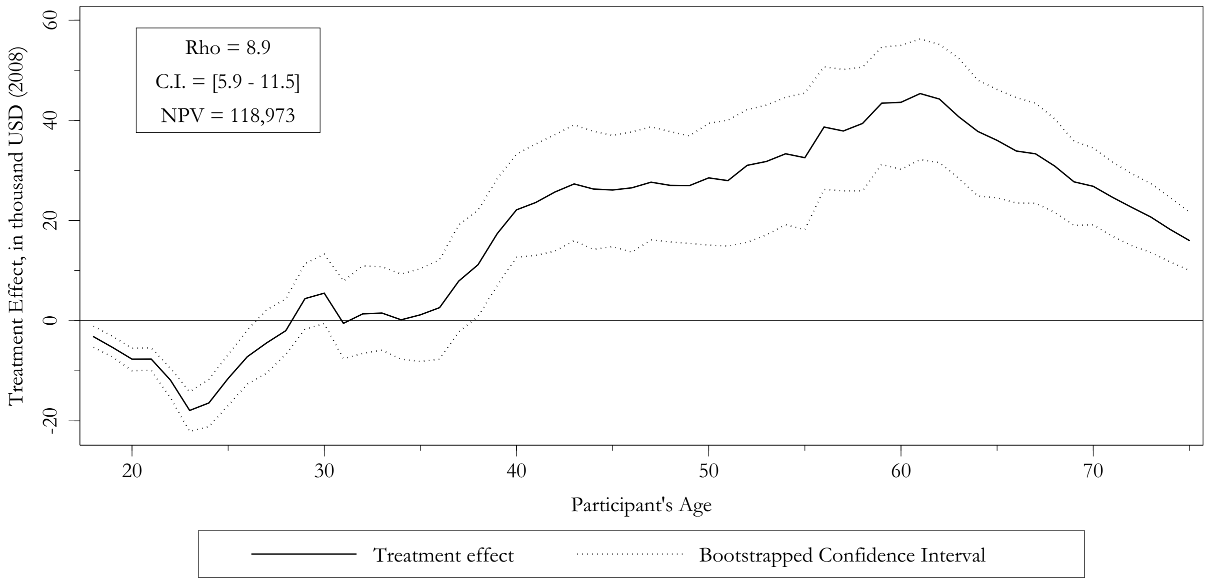
(b) Master's vs Some College



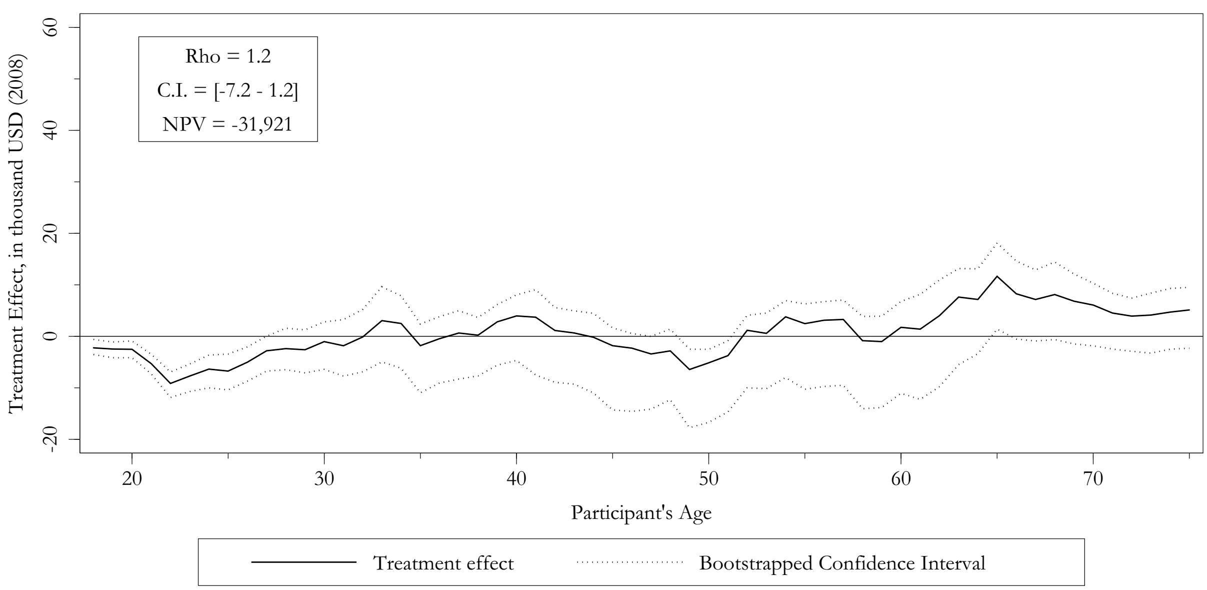
Notes: See notes to Figure 2.

Figure 5: Pairwise Treatment Effects on After-Tax Earnings, Males

(a) Doctorate vs Some College



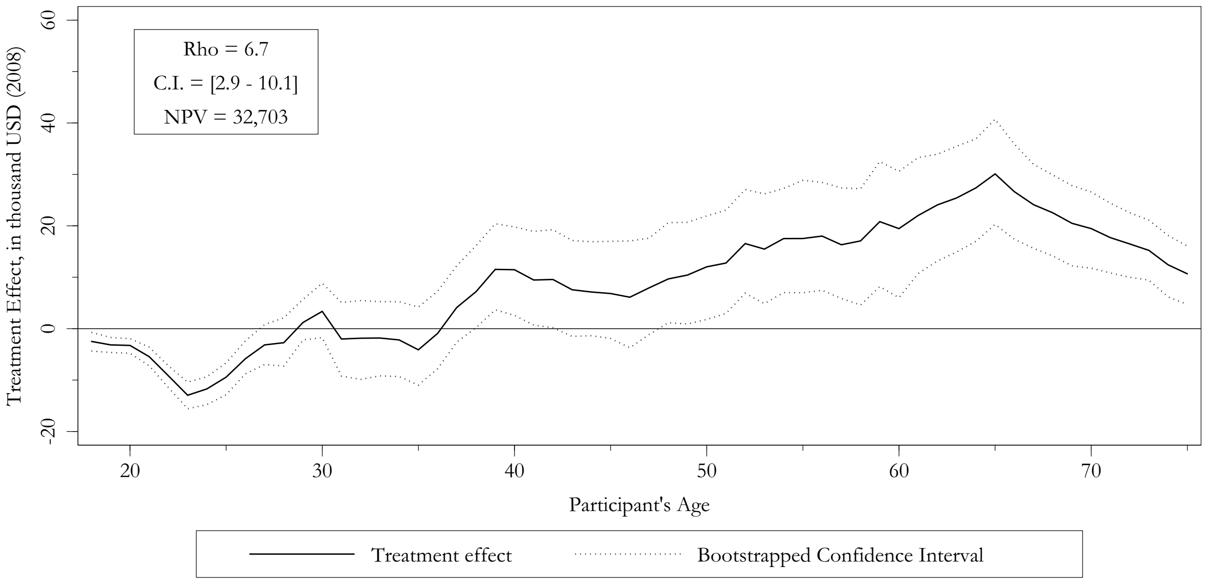
(b) Master's vs Bachelor's



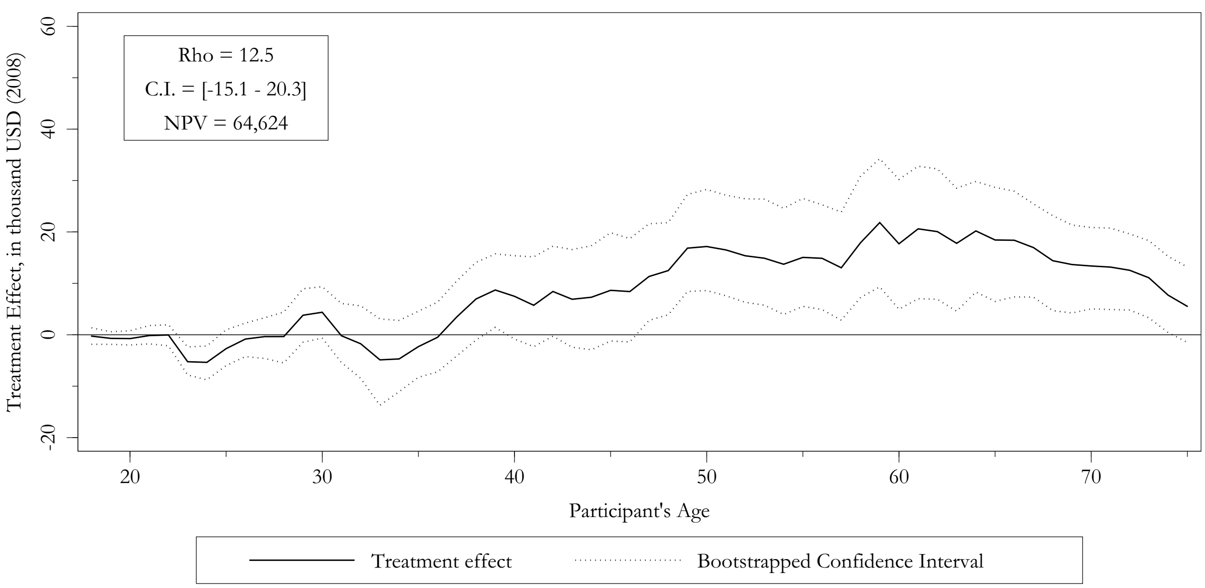
Notes: See notes to Figure 2.

Figure 6: Pairwise Treatment Effects on After-Tax Earnings, Males

(a) Doctorate vs Bachelor's



(b) Doctorate vs Master's



Notes: See notes to Figure 2.

Table 3: Internal Rates of Return and Present Values

	IRR				NPV, r=5%			
	Some Coll.	Bachelor	Master	Ph.D.	Some Coll.	Bachelor	Master	Ph.D.
High School	9.0	11.1	8.0	8.9	25,518	111,788	79,867	144,491
	[-12.6 - 16.2]	[5.5 - 15.5]	[4.5 - 11.7]	[6.4 - 11.7]				
Some College		12.7	7.6	8.9		86,270	54,349	118,973
		[0.2 - 19.6]	[3.4 - 11.6]	[5.9 - 11.5]				
Bachelor			1.2	6.7			-31,921	32,703
			[-7.0 - 4.9]	[3.0 - 10.5]				
Master				12.5				64,624
				[-15.0 - 18.4]				

Notes:
The internal rates of return represent the positive root to the polynomial $\sum_{t=T_{start}}^{T_{end}} \frac{\Delta_t}{(1+\rho)^t} = 0$, where Δ is the difference between two otherwise equal persons with high and low education, the treatment effect of education as in Figures 2 to 4. ρ was found using the mata optimizer in Stata.
The confidence intervals are basic bootstrap confidence intervals, at the 90% level, based on computations of the IRR for each of 200 bootstrap draws.
The discount rate used for the present discounted values is 5%, 3%, and 7% as indicated.
The category of "High School" includes individuals who have attended college but did not obtain a degree. The median length of college for them is 1 year of college, and the mean 1.7. In "Bachelor", persons with some graduate classes are included as well. For them, the median is 1 year of graduate school, and the mean 1.25 The tax rates are for married and single men separately, and we assume tuition was paid in full when college/university was attended (both at the undergraduate and graduate level).

Table 4: Internal Rates of Return and Present Discounted Values, Males, Simpler Estimation Methods

(a) Excluding IQ and Personality from Covariates

	NPV, r = 5%				NPV Percent greater than True NPV			
	Some Coll.	Bachelor	Master	Ph.D.	Some Coll.	Bachelor	Master	Ph.D.
High School	50,509	145,598	119,292	194,630	98%	30%	49%	35%
Some College		95,088	68,783	144,121		10%	27%	21%
Bachelor			-26,306	49,033			18%	50%
Master				75,338				17%

(b) Limited Set of Covariates

	NPV, r = 5%				NPV Percent greater than True NPV			
	Some Coll.	Bachelor	Master	Ph.D.	Some Coll.	Bachelor	Master	Ph.D.
High School	51,152	154,760	158,895	256,145	100%	38%	99%	77%
Some College		103,608	107,743	204,993		20%	98%	72%
Bachelor			4,135	101,385			113%	210%
Master				97,250				50%

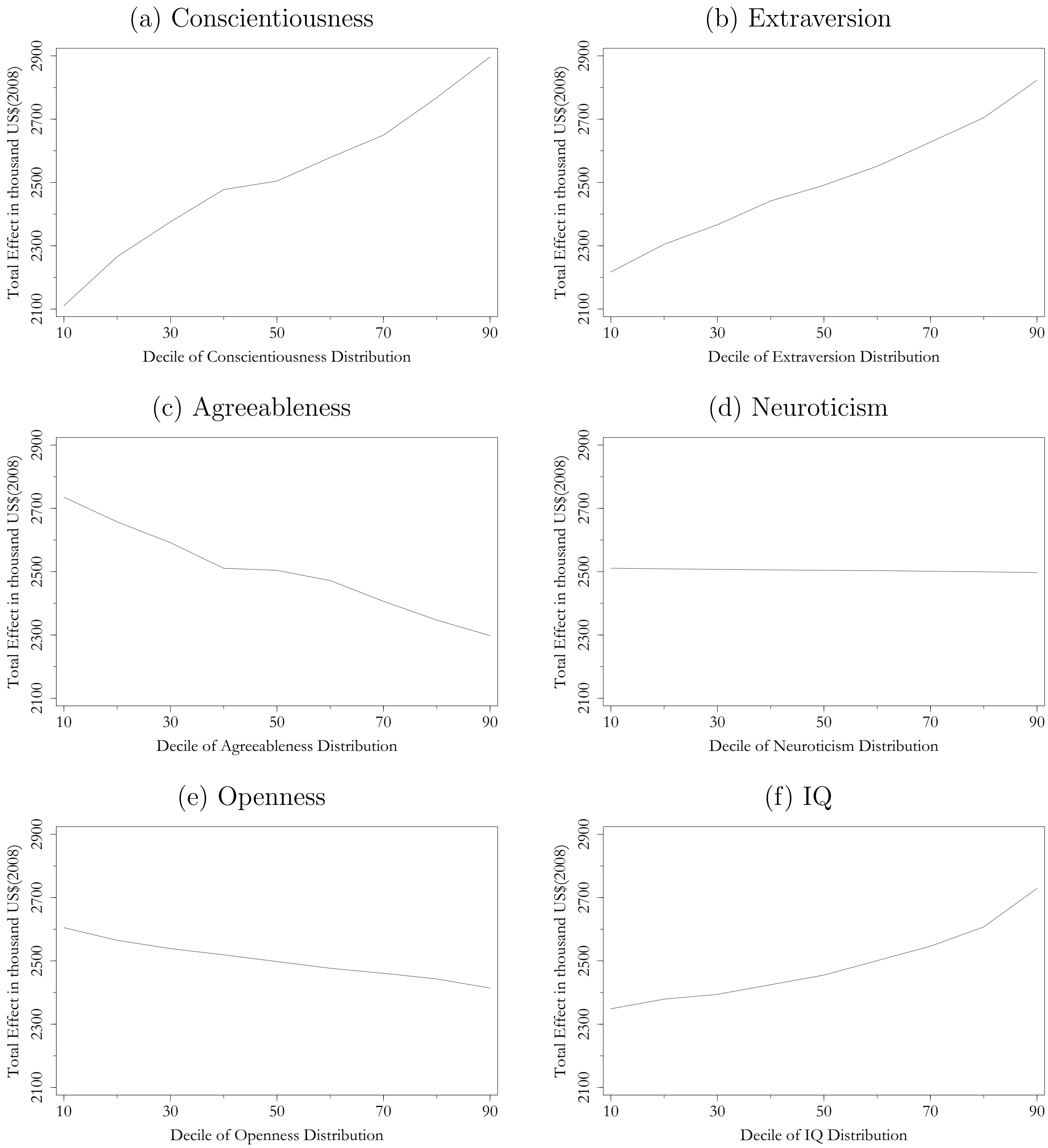
Notes:
See Notes to Table 3 for details about the methodology producing the IRR and NPV.
Part a) of this table repeats the same estimation as the preferred specification, except for dropping IQ measures and personality factors from the list of covariates.
Part b) uses a very restricted set of covariates: subject's state of birth, two cohort indicators, and parental education dummies.

Table 5: Determination of Lifetime Earnings

	Total Effect		**Total Effect with covariates**		**Direct Effect, given Education**	
	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Conscientiousness, 1940/50	412.5	(0.000)	383.8	(0.000)	313.5	(0.000)
Agreeableness, 1940/50	-288.0	(0.013)	-268.9	(0.018)	-274.3	(0.014)
Neuroticism, 1940/50	93.0	(0.419)	-8.5	(0.941)	-30.8	(0.783)
Openness, 1922	-128.0	(0.145)	-86.4	(0.325)	-110.3	(0.199)
Extraversion, 1922	347.5	(0.001)	350.6	(0.001)	354.3	(0.001)
IQ, 1922	201.5	(0.005)	162.3	(0.034)	137.1	(0.067)
High School diploma					-1,260.7	(0.000)
Some College					-1,077.0	(0.000)
College degree					-538.1	(0.003)
Graduate degree					-495.7	(0.015)
Full set of Controls	No		Yes		Yes	
Observations	591		591		591	
Adjusted R-squared	0.061		0.131		0.176	

Notes: The dependent variable is lifetime earnings in thousand US Dollars of the year 2008. It is the sum of all earnings, after tax and tuition, from age 18 to 75 for each Terman male in the estimation sample, undiscounted. For a description of the generation of the lifetime earnings, see Appendix A, and Section A.4 for the estimation sample.

Figure 7: The Total Effect of Personality and IQ on Lifetime Earnings



Notes: The predictions of lifetime earnings are based on the regression presented in Table 5. Lifetime earnings are the per-person sums of earnings from age 18 to 75, in thousand US-Dollars of the year 2008. They are not discounted. As a baseline, we used means of all covariates, and vary only the factor score in question; thus implicitly holding constant all variables including the other factor scores.

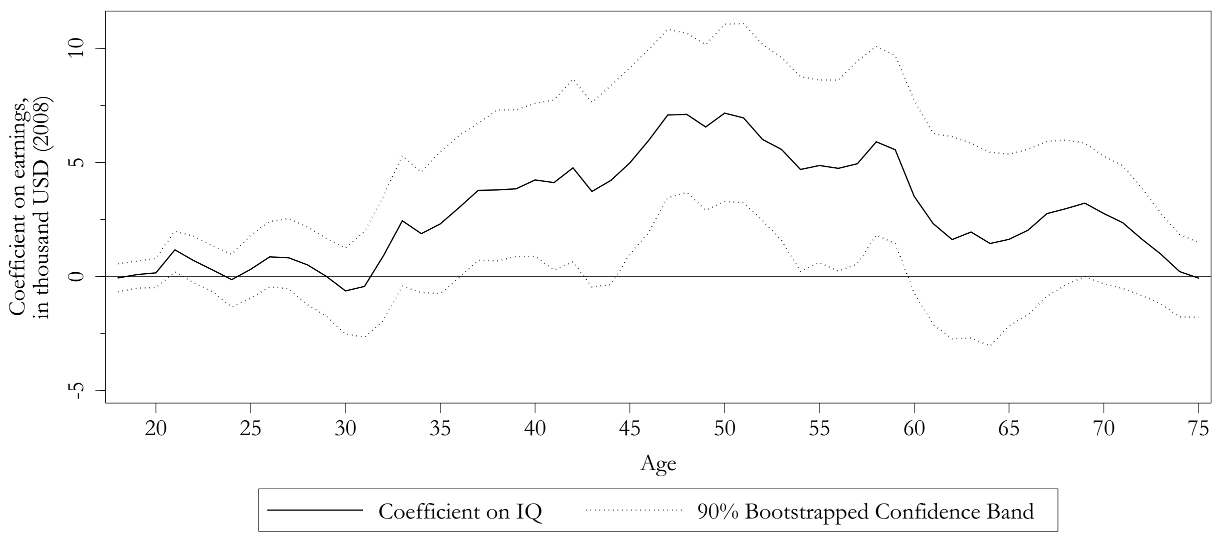
Table 6: Personality Traits in Educational Choice

	Multinomial Logit Estimation							
	Some College		Bachelor		Master		Ph.D.	
IQ	1.51	*(.115)*	**1.64**	***(.045)***	**1.84**	***(.016)***	**1.68**	***(.038)***
Conscientiousness	1.08	*(.713)*	1.21	*(.362)*	1.35	*(.192)*	**1.97**	***(.003)***
Neuroticism	**0.43**	***(.012)***	0.71	*(.254)*	0.71	*(.281)*	0.79	*(.440)*
Openness	0.13	*(.183)*	1.94	*(.656)*	1.11	*(.932)*	0.12	*(.149)*
Openness^2	1.27	*(.289)*	0.90	*(.628)*	0.95	*(.824)*	1.37	*(.160)*
Extraversion	**0.01**	***(.079)***	**0.00**	***(.032)***	**0.00**	***(.015)***	0.09	*(.383)*
Extraversion^2	**2.14**	***(.059)***	**2.35**	***(.027)***	**2.71**	***(.013)***	1.47	*(.334)*
Agreeableness	130.86	*(.174)*	1.46	*(.885)*	2.43	*(.775)*	127.01	*(.143)*
Agreeableness^2	0.51	*(.189)*	0.92	*(.835)*	0.87	*(.764)*	0.50	*(.155)*

Notes:
We present relative risk ratios on the predicted personality factor scores in a multinomial logit estimation. The baseline is high school diploma. The full list of covariates as used in the matching regressions above were also included as regressors. The regression sample is also the same as for the outcome equations (lifetime earnings).

Figure 8: Direct Effect of Personality on Earnings

(a) IQ



(b) Openness

Notes: Graphs show standardized coefficients δ from equation (2.1). The standard deviation in IQ represents 10 IQ points.

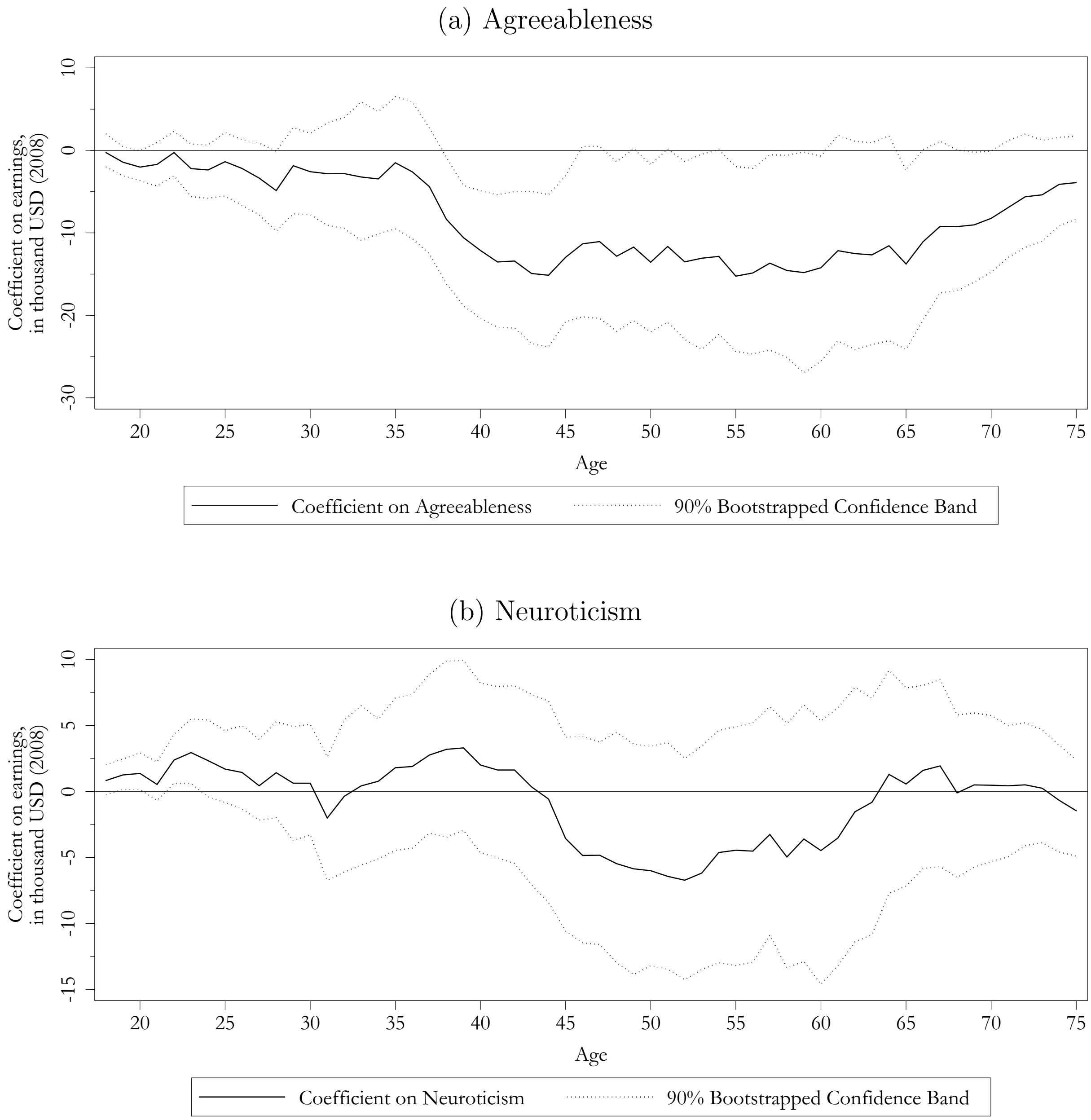
Figure 9: Direct Effect of Personality on Earnings

(a) Conscientiousness

(b) Extraversion

Notes: Graphs show standardized coefficients δ from equation (2.1).

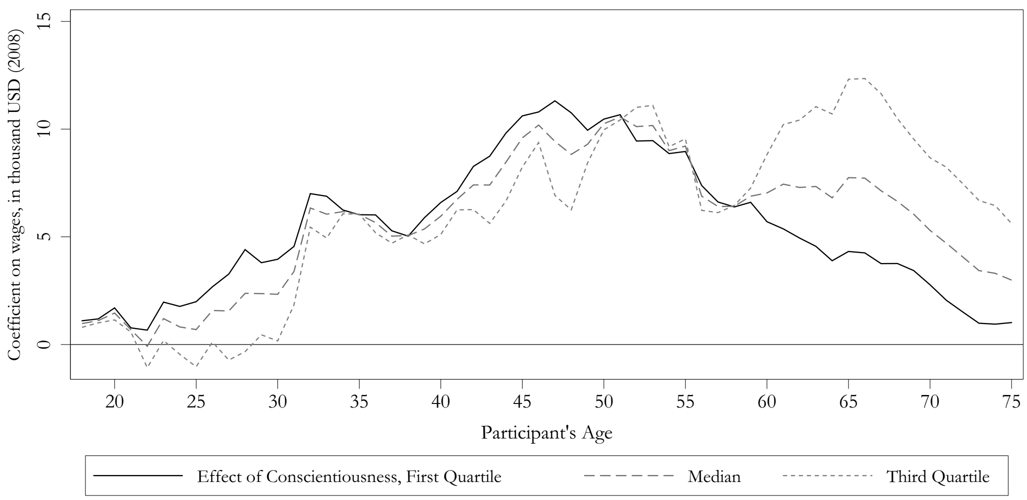
Figure 10: Direct Effect of Personality on Earnings



Notes: Graphs show standardized coefficients δ from equation (2.1).

Figure 11: Nonlinear Effects of Personality on Earnings

(a) Effect of Conscientiousness



(b) Effect of Extraversion

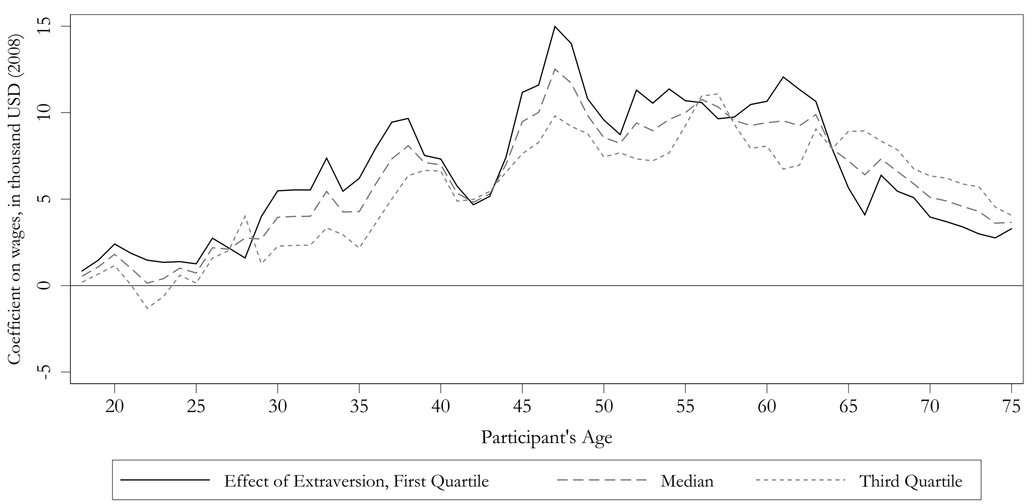
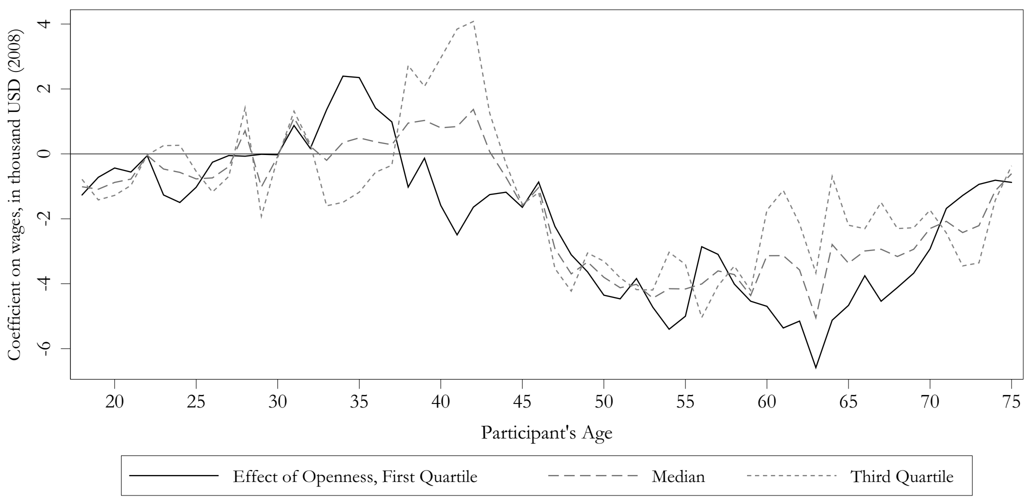


Figure 12: Nonlinear Effects of Personality on Earnings

(a) Effect of Openness



(b) Effect of Agreeableness

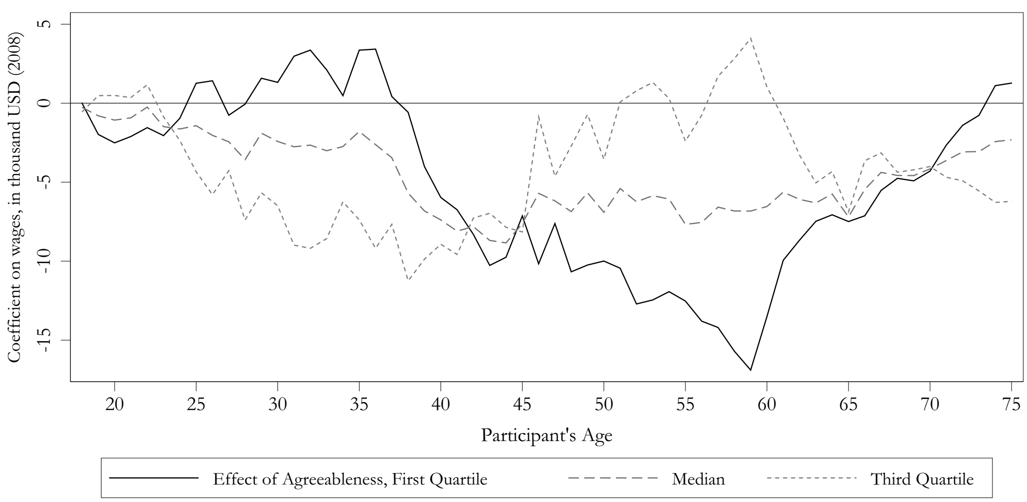
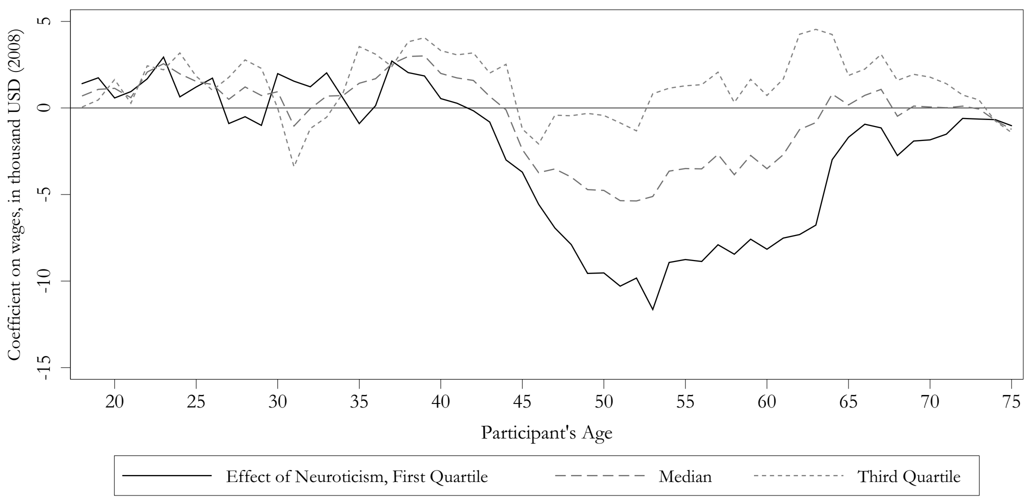


Figure 13: Nonlinear Effects of Personality on Earnings

(a) Effect of Neuroticism



(b) Effect of IQ

A Appendix A: Constructing the Earnings History

This appendix describes the data used for the analyses in this paper. From information given directly in the Terman surveys, we construct an education history, an earnings history, and a marriage history for the men and women of the Terman sample. The marriage history will serve to determine the tax-payer status as single or married. (In another paper, we also analyze the Terman females' earnings. For them, *family* earnings are much more important than own earnings). Furthermore, we match the college information given in the Terman surveys with publicly available information about tuition and fees in order to include the cost of tuition in the earnings stream (as negative numbers). The Terman data is of high quality with relatively few cases of missing information. In these few cases, we augment missing information when it is possible. For this, we apply reliable, conservative procedures detailed below. Throughout this description, we refer to the Terman study code books (Terman and Sears, 2002a,b; Terman et al., 2002a,b).

The earnings history will be described first, then the marriage history, followed by the tuition information. Finally, we will describe the sample that was used for the treatment effect computation.

A.1 Earnings History

The participants in the Terman study were born from 1900 to 1921, but since the number of subjects born before 1904 or after 1915 is relatively small, we follow the example of earlier Terman studies and do not include those subjects in the estimation sample. The overall time period of our study is thus 1904-1991, and we calculate the earnings profiles from the earliest to the latest reported paid employment.

We begin by discussing the general methods used for imputing missing earnings information. These methods apply to all periods. Some periods are special cases that require extra imputations. For such periods, we discuss our methods in the text that describes the period specifically.

In describing our construction of the income history for each year, we refer to surveys in year "19xx". Note that even if the survey is called "1936", respondents might have filled in the form in a different year. Some subjects filled it in early, other responded after the 'official year'.

A.1.1 General Methods

Imputation For every period, we compute the average (of non-missing) earnings by occupation group. We then identify persons who report an occupation but not earnings. We impute their earnings with the average of their respective occupation group. Similarly, we impute the duration of an occupation using the mean tenure of the respective occupation. Note that we impute neither the earnings nor the duration if the occupation itself is missing and the respondent did not fill in this particular survey. If, on the other hand, he filled in the survey but the occupation is coded as "'none' specified, or no [second] mention, or N.A.," we proceed as follows: If he was in schooling in the period covered by this survey, and he did not specify any occupation, we assume that he was studying full-time and did not have a paying job. If the subject did not specify an occupation but reported income and duration, we use the latter two for the earnings history. When a survey allowed several mentions for an occupation, we combine durations or incomes of all the mentions to generate the means by occupation. One exception to this mean-imputation approach concerns duration - when the subject stated that he worked "Intermittently, contract, per job, etc.," this answer is represented as missing in the data. However, it contains more information than a simple missing. Therefore, we code these occupation spells to have lasted for 3 months to represent the fact that the employment was not held with the same intensity as 'regular' occupations.

When earnings are missing, but the occupation is stated to be "housewife," "unemployed," "student," "physically or mentally incapacitated," or "retired" we impute the earnings to be zero. This is necessary in very few cases.

This general imputation method allows us to estimate earnings per age for all ages covered by the survey years. Since we only impute missing values using information from *other* individuals' earnings or durations, we can continue to proceed even if one individual has more than one type of missing information at a given age.

Piecewise Linear Interpolation Note that even if all forms were filled in perfectly by every participant, we would still have missing earnings at certain ages because of the timing of the Terman follow-up studies, which leaves gaps between some years. For example, the survey of 1955 gives us annual earnings for 1954, but the 1960 survey reports earnings only from 1956 onward. For the ages corresponding to years such as 1955, we cannot know the participants' earnings. In such cases, we use piecewise linear interpolation for the gaps between these ages in our computation of average earnings. These earnings can be considered as missing perfectly at random, since the reason for the gap lies outside the participants' motivations.

Matching Ages to Years Most of the later follow-up surveys ask explicitly about earnings *in the year 19...* To translate earnings in a given year into earnings by age, we take the participants' age at the 30th of June in a given year.

For earlier surveys, earnings are given for the "most recent occupation." We assume this refers to the participant's current occupation at the time he filled out the form. We determine the age at which the questionnaire was filled out by using the date of filling out the form (which can be before or after the year the questionnaire was officially issued). Not all respondents report a date for when they filled in the questionnaire. In these cases, we impute the age of filling out the form with the year the questionnaire was issued, the mode for filling out the form.

Whenever a subject reports a beginning age and a duration for an occupation, we fill in the earnings for this age. For example, if a job spell's duration was 24 months and the subject began the job at age 16, we assign 12 months of earnings to age 16 and 12 months to age 17. We treat all job spells alike. For questionnaires where more than one mention is possible, we sum the earnings from all job spells at a given age and obtain total earnings per age for all our subjects. Note that most of the instances where the occupations are taking place at the same age concern short-term employments at young ages.

Adjustment for Inflation We convert all nominal dollar values into US Dollars of 2008 using the US Consumer Price Index. For wages that correspond to occupations held for more than 1 year, we do not adjust the nominal values over this time span. For example, in the case of a respondent who stated in 1928 that he held an occupation for 4 years and reported his wage, we use the same wage for all four years.

A.1.2 Before 1922

The employment- and earnings-related variables collected in the 1922 survey are "paid employment" (3 mentions), "age at each employment," and "duration of each employment." The types of employment listed are jobs that we would generally expect to be filled by teenagers; this variable does not have the same code as the later occupation variable code (delivering newspapers, garden care, cleaning, baby-sitting, saleswork in stores, stenography, office work and other similar options). This variable refers to paid employment from birth up to 1922.

Methods of Calculating the Earnings Profile from Birth to 1922 The 1922 questionnaire has information about occupation and tenure, but not wages. Therefore, we have no earnings information even for the respondents who report having worked during this period. We impute wages using the income information from the 1928 form. Since the birth dates of the participants in the Terman study vary, there are many subjects who, during the 1922-1928 period, are at the same age as other subjects during the years leading up to 1922. The study's subjects are fairly homogeneous: they are all exceptionally intelligent, generally from middle-class families, and living in California. Therefore, their earnings should be expected to be comparable. We considered adjusting for productivity growth during the period, but decided against it. The occupations held by these young men are of the unskilled type only (see the description above). While we do not want to deny the considerable productivity gains in the first three decades of the 20th century (especially due to introduction of electricity), we do not believe these apply to the types of occupations we are considering here, like baby-sitting or delivering newspapers. Note that the earnings from the 1928 survey are already expressed in 2008 Dollars. We do not use all reported earnings during the 1922-1928 period, but rather the reported earnings in 1922-1924, because this smaller period is closer to the 1913-1922 period. We find the average earnings at each age for all subjects who reported employment in the 1922-1924 period. Then, we match subjects in the 1922-1924 period with subjects of the same age in the 1913-1922 period, and assign the average wage for each age from the 1922-1924 period to subjects of that age in the 1913-1922 period who report employment. Note that we take into account the tenure of the occupations from the 1922 survey in imputing earnings.

A.1.3 1923-1928

The employment- and earnings-related variables collected in the 1928 survey are "paid jobs during last 5 years" (5 mentions), "duration of each job," "age at time of each job (or beginning age)," and "compensation per week."

Methods of Calculating the Earnings Profile for 1923-1928 The 1928 survey asks the respondents to list their paid jobs during the last five years, which we use to compute earnings from 1923 to 1928. We use four variables to construct the earnings profile: the occupation category for each job, the duration of each job, the age at the beginning of each job, and the compensation per week on each job. Since tenure is expressed in months, we convert compensation per week to compensation per month. Earnings from all mentions are summed (if they accrued at the same age). To impute missing values, we use the general methods stated at the beginning of this appendix. There are a few special cases: The duration code "0" in this period indicates "no jobs mentioned or no further mention." If the respondent did list an occupation or income from that job, we recode the duration to missing, and impute it using the usual method.

A.1.4 1929-1940

The variables related to employment earnings collected in the 1936 survey are of different types. The occupation-related variables are "occupations since leaving school" (up to 4 mentions), monthly income from these occupations and the "amount of time in occupation" (which is reported in 1 year-ranges, starting from less than 1 year, going up to 7.5 years). There is also information on the "number of years of undergraduate scholarships or fellowships," "total stipend from fellowships," and "number of years of undergrad assistantships" with "total stipend from assistantships," as well as the graduate equivalents. There is also information on the "approximate total earnings as an undergraduate." The 1940 survey lists the "subject's most recent occupation", the "compensation per month" and the "number of years employed" at the most recent occupation.

Income Code in the 1936 Survey A note is in order about the income coding. For 1940, Codebook 23-24 shows that income was coded in 50-dollar increments ($50 or less, $51-$100, $101-$150, and so on).[21] Codebook 19-20, for the 1936 survey, uses 200-dollar increments instead.[22] When applied to the raw data, this code yields unbelievably high earnings. Average monthly earnings would be at around $600, which would translate to $7,200 annually. In 2008 dollars, this corresponds to *average* annual earnings of over $111,000! Remember that the average age of Terman subjects in 1936 is 26 years old. A publication by Hollingworth et al. (1940) describes the mean monthly earnings for males to be around $150 ($1,800 an-

[21] Codebook 23-24 in Terman et al. (2002a), page 58.

[22] Page 47 of Codebook 19-20 states the code for the income variables (B36072-B36075) to be the same as variable B36063, "Amount of money borrowed for college expenses."

nually).[23] Another publication, Terman (1939), describes the average annual earnings of the "A" and "C" groups, which are "roughly the upper and lower quartiles" of the boys in the sample. Terman writes that the A group, comprised of the most successful individuals, has an average salary of about $3,600 per year by age 30. The C group (least successful) earns about half of this amount. Therefore, it cannot be that the average earnings of the entire Terman group is twice as much ($7,200) than the average of the most successful group, but this is the number that would result from the 1936 income code. Clearly, using the 1936 code produces erroneous earnings measures. Instead, the 1940 code with 50-dollar increments yields the correct numbers. To confirm, see for example pages 71-72 of Hollingworth et al. (1940), where the occupations of the highest earning men, and their annual earnings, are given. The earnings range from $6,000 to $12,000 annually (in 1936 dollars). When we employ the 1940 code, the data shows the income of these individuals to be precisely these amounts.[24] There is no doubt the 1936 code should not be used, but instead the same income coding as for income from the 1940 survey.

Methods of Calculating the Earnings Profile for 1929-1940 Clearly, the earnings variables from the 1940 and 1936 surveys differ from each other: The 1940 questionnaire asks about the most recent occupation, while the 1936 asks about occupations *since leaving school*.[25] We combine our description of the 1929-1936 and the 1937-1940 periods because we use variables from both periods to construct the earnings profile leading up to 1940. The 1936 questionnaire contains information on up to four occupations held "since leaving school," the monthly earnings for each of these occupations, and the duration of each employment spell up to the nearest year. There are two challenges in constructing the earnings profiles from the 1936 survey: finding the exact years the occupation was held and knowing the precise duration. The data from 1936 does not contain information on the beginning or ending

[23]Page 71 in Chapter II, "Status of the California Gifted Group at the End of Sixteen Years," by Lewis M. Terman and Melita Oden.

[24]The 1936 code would have produced annual incomes from over $25,000 to $49,200 (1936 dollars)

[25]The phrasing is "since leaving school, what occupation have you held?"

dates for each employment spell, so we use other available information (the information on education, age at which a degree was obtained, and type of occupation) to impute the beginning and ending dates for each employment spell. For the occupation leading up to 1940, the timing is given since the variable used refers to the most recent occupation.

The code for duration of each occupation is imprecise: a duration coded as "1" indicates that the respondent has held this job for 1 to 17 months, a duration coded as "2" indicates that the respondent held this job for 18 to 29 months, and so on. We correct this coding method by almost always assuming full years, that is, we assume "1" indicates 12 months, "2" indicates 24 months, and so on. On the average, this assumption will not introduce bias.

Most recent occupation in 1940 We create the earnings profile by working backwards. First, we find the year in which the respondent filled out the 1940 form, and assign the "monthly income in the most recent occupation" to his age in this year. We then assign this income to earlier ages, for as long as the occupation was held. We only use the annual earnings information for the period between the years the respondent filled out the 1936 and 1940 forms. We do not assign earnings for years before the 1936 survey was filled out, since earnings during that period should be reported in the corresponding survey. Only if there is no information from this survey are earnings assigned to years before 1936.

There are individuals for whom we do not have any original (non-imputed) earnings information from the 1936 questionnaire at all (about 20% of the respondents). Of these 20% 35% report valid earnings in 1940, or 47% of them have imputed earnings from the 1940 form. Using the (imputed) duration of their current occupation in the 1940 form, we extrapolate these earnings back for as long as the duration, even if that means going beyond the 1936 questionnaire date. For the 7 respondents who report "7 1/2 years or more" as the duration of their occupation, we impute the true length with 8 years.

See further notes on the link between the 1936 and 1940 surveys in the paragraph "Occupations in 1940 and 1936".

Finally, we also take into account income from stipends during college or graduate school. In comparison to earnings from full-time occupations, these earnings are low.

Occupations in 1940 and 1936 There are many respondents who state in the 1940 survey that they have been holding the same job for four or more years. We know exactly how many years ago the respondents filled in the previous 1936 survey (recall that the difference is not necessarily four years, but can be shorter or longer depending on the year in which the survey was filled in). If the duration of the job from the 1940 survey exceeds the gap between the surveys, he must have been working in the same job in the year in which he responded to the 1936 survey. Then we should also find the same occupation type listed in the 1936 survey. The difference between the tenure reported in 1940 and in 1936 should not be smaller than the gap in years between the surveys.[26] Then it is certain that this particular occupation was held since the 1936 survey through to the 1940 survey. In fact, of the four occupations mentioned in the 1936 survey, this one must have been held last, and at the time of the interview. For 127 survey participants, these conditions apply.[27]

1936 Occupations For the "occupations since leaving school" listed in the 1936 survey, we do not know the start/end dates. Even if they are chronologically ordered,[28] they could have been held simultaneously or sequentially. In order to use the information in this occupation variable, we must find the respondents' age at the time of leaving school for the last time. We use the variables on schooling history from both the 1936 and 1940 forms to generate a dummy variable that reports whether or not the individual was in school/university at every age. We find the highest schooling age by using the latest date the respondent left either of the three potential colleges or grad schools listed. Note that schooling histories are not

[26]Note that for respondents who state "8 years or more" for their tenure, this difference in tenures could be 8-tenure from 1936 or greater.

[27]Note that for this identification we only use the original data, not imputed values. In all of the identified cases, the occupation carried through was the last mention.

[28]How can we say this? All of the occupations from 1936 matching the 1940 occupation are the last mention. Furthermore, the vast majority of later mentions have higher earnings than earlier mentions.

always made up of sequential spells of education, but can have interruptions. For example, a respondent who finished high school at 17 but only entered college at 19. We identify these breaks and the date (by year) when the respondent left school, and we number them. We assume again that the breaks are exactly one year long, unless we have a correspondence to other information that makes us certain that it was less or more.

The form contains information on the monthly income from each of four occupations, but we must find the exact duration of each occupation in order to construct the annual income. As before, we assume "1" indicates 12 months, "2" indicates 24 months, and so on. However, sometimes our information shows us that the duration of the spell coded as "1" was less than 12 months. For example, there are several cases of the following type of respondent: someone who just finished his last year of schooling in the year of filling out the form and states that he has worked for "1" in an occupation. In this case we know that he has worked less than one year, so we assume he worked for 6 months. This assumption should, again, yield unbiased duration length on average.

Next, we need to know the ages at which respondents held each employment. The first and most important mechanism of attributing incomes to ages is the length of the occupations and the number of years since the respondent left his last educational institution, as well as the duration of the breaks. We always assume that occupations are held in consecutive order. We then assume that respondents worked during the (full) years in which they were not attending school. Based on this assumption, we verify whether the occupations could have been held during the years since the maximum schooling was obtained, or during the breaks between schooling spells.

The "professional status" of a respondent's occupation is used to verify whether he could have held an occupation at a certain age. Notably, if an occupation is "professional", we compare the age at which the first degree (B.A. or higher) was obtained to the age when a break began. This helps us determine whether we can assign a job spell to this particular break. For example, assume someone has just finished his last year of schooling in the year in which he filled in the 1936 form. We also have the information that at some time, he worked as an engineer for three years. Say he discontinued his schooling twice, for three years each. If he obtained his first degree at age 23, and his first break started when he was 20, we know that he cannot have worked as an engineer in this break, because he did not possess the required qualification yet. In this sense, we will use the combined information on the age of the break, the age of the degree, and professional status of the occupation, to exclude certain possibilities or cases for the timing of job spells.

In the "professional" category of occupations, we include almost all occupations in the category "Professional" in Codebook 2:

> Member of university or degree-granting college faculty: teaching, administration, research; architect (includes naval architect but not landscape); social scientist (includes economist, psychologist); political scientist, diplomatic service; social welfare worker; probation officer; chemist (includes chemical engineer); physicist n.e.c.; clergyman; engineer; lawyer; physician; optometrist; psychiatrist; biological or physical scientist except chemist or physicist (e.g., biologist, geologist, paleontologist, etc.); mathematician; Army and Navy officers, Air Corps officers; intern; resident; surgical fellow; research fellow (other than medicine).

If the occupation durations exceed the length of time since leaving school, or the length of absence from school, we start in the year of the survey and attribute earnings of all occupations mentioned consecutively into the preceding years, even if the respondent was attending school in one of those years.

Duration of Spells The preceding principles concerned the timing of the occupations. The most important rule is to attribute as many of the occupations to the period after leaving school as possible, because the question asks explicitly about occupations since leaving school. While we usually take the durations to be full years as described, we take into consideration the time since the respondent left school. If the sum of the durations of the up to four mentions is exactly one year longer than the time since leaving school, we adjust the "1" lengths downward, taking them to be 6 months only instead of 12 months. This is also true if the person just finished school (the age of leaving school is equal to the age at filling out the form), where we assume that he worked 6 months.

For a case-by-case discussion of the different occupation mentions and possible combinations of the length of breaks from school and duration of occupations, please contact the authors directly.

A.1.5 1941-1959

1941-1944 The employment- and earnings-related variables collected in the 1945 survey are very straightforward: "occupation in 1941, 1942, 1943, 1944", and "income in 1941, 1942, 1943, 1944."

The directly reported yearly earned income (in hundreds of dollars) and occupation type in the years 1941-1944 is used as is. For respondents who report an occupation but no income in that year, we impute using the mean by occupation type.

1946-1949 The 1950 survey corresponds closely to the 1945 survey, i.e. it has information on "occupation in each year from 1946 to 1949" and "earned income in each year from 1946 to 1949"(in rounded hundreds of dollars). We use the same methods for this period as we used for the 1941-1945 period.

1954 The 1955 survey lists the subject's occupation and his earned income in 1954 (in rounded hundreds of dollars). We use linear piecewise interpolation to fill the gap from 1950 to 1953, where there was no earnings information from the surveys.

1956-1959 The employment- and earnings-related variables collected in the 1960 survey correspond again to the 1945 survey, "occupation in each year from 1956-1959" and "earned income in 1956, 1957, 1958, 1959" (in rounded thousand dollars). We use the same methods for this period as we used in the 1941-1944 period.

A.1.6 1960-1972

The survey in 1972 gives us the "subject's income in 1970, 1971," and average income for 1960-1969 (all in rounded thousands of dollars). The phrasing of the questions was "What were your total *earnings from your work*? Do not include royalties, pensions, etc. from work in prior years." For the 1960-1969 average income, respondents were told, "In column (a), please give representative annual figures for the decade 1960-1969." We also know the subject's occupation, and former occupation if not currently employed. Also available are the "amount of time spent on the job" (essentially full-time to much less than half-time, or retired), "division of subject's time: income-producing work, 1960-65, 1966-1972." This latter variable enters a summary-variable in the 1982 survey about the time devoted to full-time work, which we will also use.

Even though we have variables on total earnings for 1970 and 1971, the occupation information is available only for the year 1972. Therefore, we impute missing values with the usual procedure, but using the 1972 occupation for generating occupation-averages (full occupation code, not the simplified scale), since the occupations in 1970 and 1971 are not available.

For the 1960-1969 period, we take the respondents' answers at face value. Since the question specifically asks for a representative figure, we use the amount given for all years in the decade 1960-1969.

A.1.7 1976

The employment- and earnings-related variables used from the 1977 survey are: "subject's earned income in 1976"(in rounded hundreds of dollars), "occupation" and "percent of full-time work at each age, from age 50 to age 75."

Even though data is available on social security payments, other pensions, or investment income, we do not include this in the earnings variable for 1967. We use only earned income. Especially including social security payments as part of earned income in 1976 would lead to double-counting. We use the same method of interpolation for the earlier years in this period as we used in the 1951-1955 period.

A.1.8 1981, 1986, 1991

The 1977 survey is the last survey to ask the Terman Study participants for their income. The 1982 questionnaire asks for the "percent of full-time work annually" from 1975 to 1981. The 1986 and 1991 surveys ask for the "percent of time devoted to different activities, occupational work for pay." Clearly, when none of the respondent's time is devoted to work for pay, his earnings will be zero.

We use two types of extrapolation to calculate earnings in the last years of the Terman study. The first uses the fact that participants were born in different years, so in 1977 some are older than others. We can extrapolate earnings from 1977 to 1985, using earnings of those who are older to impute earnings of the younger participants. We also use the information from the work-effort variable ("percent of full-time work"). For imputation, we proceed by birth cohort, and each time use the two closest birth cohorts available. For those two cohorts that are older than the cohort we wish to impute, we regress their earnings in 1977 on the previous years' earnings, controlling for the amount of full-time work done (in 5 categories). We then predict all respondents' earnings for this year, and use the extrapolation to impute earnings for the younger cohorts. For example, if we wish to extrapolate earnings of persons who were 69 years old in 1977, we start by using earnings of persons who were 70 and 71 in 1977. We regress their earnings at 70 on their earnings at age 69, controlling for work effort. We use the resulting prediction to extrapolate the earnings of those who were 69 in 1977 to their imputed earnings at age 70 in 1978. Next, to extrapolate their earnings up to age 71, we take those who were 71 and 72 in 1977, and repeat. This extrapolation should not introduce significant bias, because the cohorts are relatively close, implying that year-effects are unlikely to dominate the true earnings distribution. This extrapolation ensures earnings data available until the age of 72 for all males.

We next extrapolate further using the "percent of full-time work annually", when the percent of the time worked is not zero. If a respondent's reported earnings in 1967 were zero, but some time was devoted to work in that year, we assume that if his work effort in later years did not exceed that of 1967, then his earnings during those later years were also zero. If his work effort in later years is equal to the work effort in 1967 and his earnings in 1967 were not zero, we impute subsequent earnings with the earnings of 1967. Otherwise, we impute them with the proportional earnings of 1967, using the ratio of work efforts (current percentage relative to 1967). We use the same procedure using the 1986 and 1991 "work for pay" variables, to impute earnings in those years. Between these surveys, we use piecewise linear interpolation since no earnings information is available.

We also assume that subjects who are retired in 1977 — with zero earnings and zero work effort in the 1986 and 1991 surveys — remain retired throughout this period. Given the advanced age of the sample in 1991, if a respondent has zero earnings in 1991, we extrapolate this number out.

A.2 Taxes

All tax brackets and rates used here are those specified in IRS Form US 1040. We obtained the U.S. Federal Individual Income Tax Rates History from the Tax Foundation at http://www.taxfoundation.org/publications/show/151.html.

The baseline earnings histories used for the results in Sections 3 and 4 apply tax rates for married persons to all individuals for whom the current marital status is "married." For all other persons, including those for where the current marital status can not be determined, we apply the tax rates for singles.

A.3 Tuition Payments

In order to properly take tuition into account, we must first build the education history from the Terman surveys. This history identifies, among other things, the university, college, or other institution of higher education attended by the Terman participants, in which years they were enrolled, and what types of degrees they received. Based on this information, we search for the actual tuition charged at the time by these institutions.

A.3.1 Education History

The higher education histories are generated using information from the 1936 and 1940 surveys, as well as the education file from 1991.

The two surveys give a list of up to three colleges and three graduate schools attended, the duration of attendance, the year in which the subject left this institution, the year in which he received a degree, and the type of degree received (B.A., M.A., etc.). The education file in 1991 lists the degrees obtained and the year in which the degree was received. and the name of the institution.

In the majority of cases, the information from the two surveys matches exactly. In such cases, we generate a variable that specifies the name and type of institution attended (none, undergraduate, or graduate) at each age.

When there is different information in the two surveys, we analyze all cases individually. In most cases, the information difference reflects a change between 1936 and 1940 in educational attainment (for example, the respondent began college in 1937) or ordering (for example, the college listed first in 1936 was listed second in 1940). In these cases, we use the information from both surveys. In a few cases, the cause of differing information cannot be determined. In such instances, we only use the survey that concurs with the 1991 education file.

Imputation Imputation is necessary when the college name and type is listed, but not the duration of attendance. We then impute the duration of attendance with the median of the corresponding education level (for example, persons with "some college" attended college on average for one year). If the age of entering college is missing, we assume that the degree was pursued immediately following the previous education record. For example, if a person went to college but is missing the date of enrollment, we assume that he/she started college in the year after graduating from high school.

undergraduate and graduate schools for Terman males and females by year. Attendance is concentrated roughly in the years between 1920 and 1950.

Types of Programs We further try to distinguish the types of graduate school programs attended by Terman participants, as the surveys do not distinguish between general or professional graduate programs. We use occupational information in order to identify those who went to professional schools, such as law, medical, or pharmaceutical schools. Occupation variables are available in 1940-1944, 1946-1949, 1954, 1956-1959, 1972 and 1977. If the individual was recorded as "lawyer" in any of the above-mentioned surveys, we determine that he/she is a "lawyer." If the individual was recorded as "physician," "surgeon," "dentist," etc., we determine that he/she was a "doctor." If the individual was recorded to be working as a "pharmacist" or some miscellaneous medical fields in the above-mentioned surveys, we determine that he/she was a "pharmacist." Based on such classifications there are 31 doctors, 67 lawyers, and two pharmacists in the sample of males, and seven doctors, 20 lawyers, and two pharmacists in the sample of females.

Since practicing in these professions requires the corresponding professional degree, we conclude that these individuals must have attended a "professional" graduate school such as law school, medical school, or pharmacy school.

Therefore, for each college or university, there can be five types of programs for each year: "undergraduate," "graduate," "law," "medicine," or "pharmacy."

Educational Attainment For any categorization by education in our analyses we use the educational attainment as listed in the 1991 education file. This represents the final attainment, in the sense that it includes any education the respondent will ever obtain in their lifetime. This attainment is given by Terman researchers as an "omnibus measure," and it is supposed to represent the best available information. In very few cases, we are not been able to verify this educational attainment from the separate variables and we adjust accordingly.

Four women are listed in the category "high school only," but it shows on the 1936 and 1940 surveys that they attended college. For these women, we change their final educational attainment to "some college" (ID-numbers 621, 634, 1007, 1511). In all of the main analyses these two education groups are subsumed into one, so this change is immaterial.

A similar situation occurrs for two males (ID-numbers 592, 793). One man is in the "B.A." category in the Education File, but lists graduate school in 1936 and 1940. We change his educational attainment to "some graduate school" (ID-number 999).

A.3.2 Direct Sources for Tuition Information, between 1920 and 1950

Most Terman subjects attended institutes of higher education directly after high school, so tuition is needed for the years 1920 through 1950.

We use two major series of books for tuition information during these years: "American Colleges and Universities" (de Gruyter, W., ed., 1948) and "College Blue Book" (Hurt, H., ed., 1949). The former series is available in the years 1936, 1940, and 1948, whereas the latter series is available in the years 1923, 1928, 1933, 1939, and 1949.

For our purpose of measuring the cost of higher education, it is sufficient to know how much the individuals had to pay for schooling, including tuition and required general fees. We exclude room and board expenses from our analysis since people in the work force must also pay similar expenses. The "College Blue Book" provides tuition and general fees separately. The general fees were generally only for non-optional items to be paid by the students. We combine tuition (non-resident tuition, where applicable) and general fees to get the tuition rate. The "American Colleges," however, do not always list a separate expense item called "general fees," but rather give more detailed lists of different types of fees. Such items include "registration fees" and "student activities fees." Whenever it was not clear whether an expense item is required (such as a gym fee), we compare the total amount of fees computed to the number listed in the last available "College Blue Book."

California Schools Since the Terman subjects were geographically located in California, a majority of them attended universities and colleges in California, especially the University of California, Berkeley and Stanford University. Generally, whenever the tuition differs between residents and non-residents, we use resident rates for schools in California, and non-resident rates for schools outside of California.

The website for Stanford University provides some information about the historical tuition by decade.[29] University of California, Berkeley provides rates for its law school in 1920 and 1930.[30] We use this information in addition to the main source books.

Junior Colleges, Teachers' Colleges, California State Colleges, and Theological Schools Junior colleges, teachers' colleges, California state colleges, and theological schools were mostly two-year programs designed to provide a minimal or specific form of higher education. Their tuition was much lower than at private four-year colleges, and in most cases, such institutions were free for residents.

Furthermore, these universities (and Claremont College) only had general graduate programs and did not offer professional degrees, such as law or medical degrees. Therefore, we use the tuition for general graduate programs in these schools, even if the respondent is listed as a lawyer, doctor, or pharmacist.

[29] http://www.stanford.edu/about/facts/chron.html

[30] http://www.law.berkeley.edu/alumni/gift/tuition.html

Military Schools Terman participants attended two national military schools: Annapolis and West Point. During the relevant years, no tuition was charged and students received stipends from the government. In 1939, the stipends were $780 dollars per year for Annapolis and $1,072 dollars per year for West Point (Hurt, H., ed., 1949). We impute (-$780) dollars as tuition for Annapolis and (-$1,072) for West Point.

Foreign Institutions Tuition for Oxford University and the University of Cambridge in the United Kingdom are available in the annual series "Yearbook of the Universities of the Empire," published by the Universities Bureau of the British Empire. We use the tuition listed in 1931, 1932, 1933, 1934, and 1937 and apply the exchange rates available at the website of Measuring Worth.[31]

Tuition for the University of Berlin is obtained from "Gebührenordnung für das Wintersemester 1932/1933 des Preußischen Ministers für Wissenschaft, Kultur und Volksbildung," cited in von Lösch (1999).

An additional university category is labeled "Other foreign institutions." For this category, we impute the average tuition by year for the University of Berlin, Oxford University, and University of Cambridge.

A.3.3 Imputation of Tuition

Stripps College Stripps College was a college in California which charged a lump-sum for tuition, general fees, and room and board.[32] In order to estimate the tuition and general fees, we use the "Low student expense" from the "Blue Books," which accounts for the estimated minimum student expense for each university. We calculate the average proportion of tuition and general fees in the low student expense for universities in California (excluding junior colleges, teachers' colleges, California state colleges, and theological schools), and apply the proportion to the lump-sum amount owed to Stripps College. In 1939, this amount was $394

[31] http://www.measuringworth.org/exchangepound/

[32] In 1939, this amount was $1200 dollars per year (Hurt, H., ed., 1949).

dollars.

State Universities Outside California Terman subjects who attended universities outside of the state of California usually attended prestigious, private four-year universities, such as Ivy League universities or large four-year state universities. We assume again that Terman subjects are residents of California and apply non-resident tuition rates.

For the category "Out of state college or university, not listed, such as Bryn Mawr," we impute the average non-resident tuition of all other state colleges and universities, including Iowa State University, Ohio State University, University of Oregon, University of Arizona, University of Illinois, and University of Michigan.

Missing School Information When the name of the college is missing, we impute the tuition with the average tuition of all Terman subjects by gender and year.

Missing Tuition for the Particular Program As described above, we obtain tuition information for five types of programs: "undergraduate," "graduate," "law," "medicine," and "pharmacy." Whenever the tuition for one particular type of program in a year is missing, we look at the tuition in other years at the same university. We compute the average ratio of tuition for the program in question to the tuition for another program, and use the ratio to impute the tuition of the first. For example, assume we have information on undergraduate tuition at Princeton from 1930 to 1935, but only information on graduate tuition at Princeton up to 1934. To impute graduate tuition in 1935, we would calculate the average ratio of graduate to undergraduate tuition from 1930 to 1934, and multiply this ratio by undergraduate tuition in 1935.

Tuition in Later Years Snyder (1993) provides the total enrollment in four-year private and public universities and colleges for every year since 1920. We also have access to average tuition for private and public universities and colleges from the American Council on Education, ed. (1979) (in 1939, 1947, 1951 and 1955) and the **?** (in 1981-1984). We calculate the weighted average of tuition for private and public institutions by enrollment. Average tuition of colleges in 1949 is provided in Conrad (1956), which we use directly.

To compare the historic tuition rates to current levels, we gather tuition rates for the year 2008. At the undergraduate level, tuition and required fees are taken from the "College Blue Book" (Hurt, H., ed., 1949). The graduate school tuition rates are not available, and thus we impute them using the average ratios between undergraduate and graduate rates from earlier years. For professional schools, including medical and law schools, the tuition rates are available at the US News website. We took 2008-2009 tuition for medical schools[33] and 2009-2010 tuition for law schools.[34] 2008 tuition rates for professional schools at the University of Pennsylvania and the University of California, Berkeley are taken from the university websites.

beginfigure

A.4 Sample for Treatment Effect Computation

We include all males and females (856 and 672; 1,528 total) in our creation of the wage history. For the treatment effect computation, however, we use a more homogenous sample. There are two parts to our sample selection: the first one follows a similar approach as Martin et al. (2007). The second one is concerned with the wage histories and treatment effect analysis specifically.

First, we follow standard procedure in the literature by excluding the study's youngest and oldest students, since their selection into the sample was non-standard. We retain only those respondents born between 1904 and 1915. This is the case for 766 men. Next, we exclude individuals who never participated or dropped out of the study before 1940 (20 males). The study participants who are missing both parents' and teachers' socio-emotional

[33] http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-medical-schools/research-rankings/

[34] http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-law-schools/rankings

trait variable cannot be included either (25 males). We also exclude study participants who did not attend school, and those who did not obtain a high school diploma. These subjects did not reach a higher schooling level because they died in their early teenage years. Including them in the treatment effect computation would yield misleading results (6 males). In order to avoid reverse causality, we also exclude participants with rare genetic diseases (for example, Chorea or Hodgkins Disease; 4 males). Finally, there are four males for whom the education information is missing.

For the treatment effect analysis, we want to ensure that we have a homogenous sample. 60 males attrited from the sample before 1970, or it was impossible to establish contact. We keep only white participants - excluding the children born on Indian Reserves, and those whose parents are potentially non-white (from South Africa, the Caribbean, and China). Their experience in school and the labor market might or might not have been affected for example by prejudices, differential treatment, and the like, we do not want to make the matching assumption for them. Since ex post it is impossible to verify whether they were indeed just as the other participants in the sample, we judge it to be most conservative to not include them in the computation of the treatment effect. Discarding the above cases leaves us with 634 males in the sample. 608 males have the full list of covariates for our treatment effect analysis (non-missing).

In order to follow one group from the beginning of the study until the end, we want to include only individuals with complete earnings histories. Therefore, we add the following condition: We select only subjects with at most 10 years of missing income data. This leaves us with 575 males.

B Appendix B: Additional Material

B.1 Functional Form Comparison for Treatment Effect Estimation

Here, we briefly present three different possibilities for dealing with the functional form and illustrate with estimates for each. The successive methods will be increasingly restrictive. We analyze the treatment effect for males from obtaining a doctoral degree in comparison to a bachelor's degree.[35]

Local Linear Matching The first estimation method we propose does not impose a functional form on the outcome equation. Here, matching is implemented through a generalized form of kernel matching on the propensity score: local linear regression matching, due to Heckman et al. (1997). Figure B-2 presents the treatment effects from all three estimation methods by age. Note already that there is almost no difference between the estimated curves. Figure B-1 allows us to verify assumption (M-2). We see that the propensity score does not span the *entire* range from zero to one, but comes close.

Parametric Model The next method assumes a particular functional form for the $\boldsymbol{\mu_k}(.)$:

$$\boldsymbol{\mu_1}(X_t, \theta) = X_t\beta_1 + \theta\delta_1,$$
$$\boldsymbol{\mu_0}(X_t, \theta) = X_t\beta_0 + \theta\delta_0;$$

where the treatment effect becomes

$$\Delta_t = X_t\beta_1 + \theta\delta_1 - X_t\beta_0 - \theta\delta_0 + \varepsilon_{1,t} - \varepsilon_{0,t}, \tag{1}$$

[35] We chose this example because these are the largest groups in the Terman sample, and we thus have one of the largest possible sub-samples by focusing on these education levels. Furthermore, this comparison is of great interest because in other datasets, it is often plagued by lack of observations. The results in this section, however, do not depend on which education pairs we use. Finally, note that the category "bachelor's degree" includes persons who attended some courses at the graduate level but did not obtain an advanced degree.

and one can regress the observed Y on the observed X, latent θ, and an interaction of these with the treatment indicator:

$$
\begin{aligned}
Y_t &= X_t\beta_0 + \theta\delta_0 + D\left(X_t\beta_1 + \theta\delta_1 - X_t\beta_0 - \theta\delta_0\right) + e \\
&= X_t\beta_0 + \theta\delta_0 + DX_t\left(\beta_1 - \beta_0\right) + D\theta\left(\delta_1 - \delta_0\right) + e.
\end{aligned}
$$

Alternatively, β_1 and β_0 can be estimated in separate regressions. To obtain the treatment effect, the difference between the coefficients is multiplied with the relevant sample means of the covariates (here, we always present average treatment effects, therefore we apply the overall sample mean). The second line in Figure B-2 is the result from this treatment effect computation.

The two methods just presented have the drawback of being relatively data-hungry or more sensitive to changes in specification when the number of observations is small. We are interested in computing pairwise rates of return, therefore the sample size for these estimations will necessarily be only a fraction of the overall sample size. Therefore, we explore the third method, which makes stronger parametric assumptions but is more stable than the previous two methods.

Common Coefficient Model If we assume the *same* effect of covariates in both outcomes, such that

$$
\begin{aligned}
\mu_1\left(X_t, \theta\right) &= c_{1,t} + X_t\beta + \theta\delta, \\
\mu_0\left(X_t, \theta\right) &= c_{0,t} + X_t\beta + \theta\delta.
\end{aligned}
$$

The treatment effect is now simply

$$
\Delta_t = c_{1,t} - c_{0,t} + \varepsilon_{1,t} - \varepsilon_{0,t}. \tag{2}
$$

Estimation of the treatment effect will consist only of regressing the observed Y on the observed X, latent θ, and the treatment indicator:

$$Y_t = X_t\beta_0 + \theta\delta_0 + D\overline{\Delta}_t + e.$$

Note that this particular form of matching is linear parametric matching, or OLS as a special case of matching. "In OLS, linear functional forms are maintained as exact r epresentations or valid approximations"(Heckman and Vytlacil, 2007). Figure B-2 shows clearly that all three estimation methods produce very similar results. There is no evidence that using the parametric approach with a common coefficient is too restrictive for the purpose of this paper. It has the advantage of being more stable and appropriate for smaller subsamples. Another argument in favor of the OLS method is that it allows an easy correction for the measurement error introduced by using factor scores rather than the true factors.[36] We therefore adhere to the common coefficient model for all estimations.

B.2 IRR versus NPV

To see how different tuition payment streams affect the IRR, consider the following example: Life is divided into 2 periods, the first one for schooling and the second one for work. Assume that earnings are zero in period 1 and in period Y. We will discuss two different payment scenarios: Under scenario 1, all of tuition T is paid during the school period. In scenario 2, half of the tuition is paid during schooling, and during the working period, the second half is paid back including the interest payments.

Note first that the net present value (NPV) remains constant in the two scenarios, as long as the cost of funding (the interest rate to be paid) is used to discount the payments back to the present.

The IRR is a ρ such that the NPV is zero:

[36] For a description of the correction method, see Heckman et al. (2010), Web Appendix G.

$$\begin{aligned} 0 &= -T + Y \frac{1}{1+\rho_1}, \\ \rho_1 &= \frac{Y}{T} - 1. \end{aligned}$$

The IRR for the second payment version is such that

$$\begin{aligned} 0 &= -\frac{1}{2}T - \frac{1}{2}T\left(1+r\right)\frac{1}{1+\rho_2} + Y\frac{1}{1+\rho_2}, \\ \rho_2 &= 2\frac{Y}{T} - (1+r) - 1. \end{aligned}$$

Clearly, the two are not equal unless

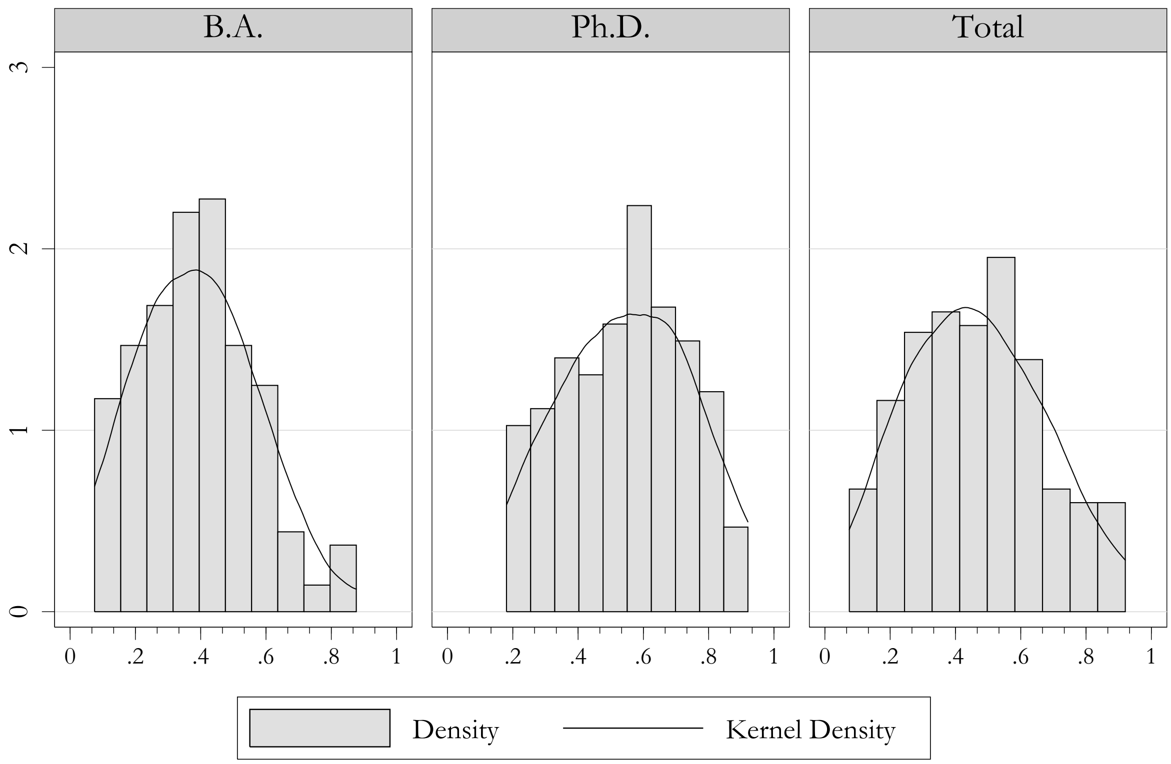
$$Y = T\left(1+r\right).$$

How do the two IRRs relate to each other? We know that

$$\begin{aligned} \rho_1 &> \rho_2 \\ &\text{only if} \\ \frac{Y}{T} &> 2\frac{Y}{T} - (1+r) \\ T\left(1+r\right) &> Y \end{aligned}$$

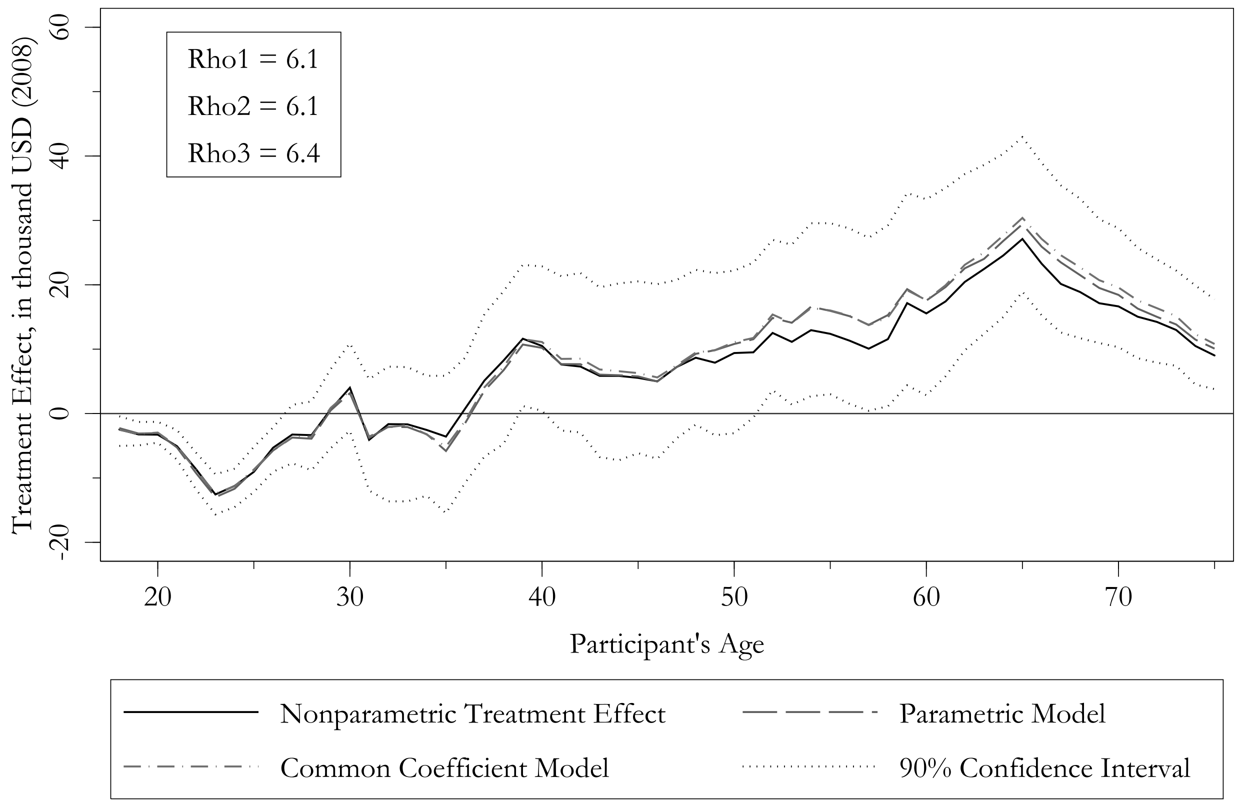
This case will generally not be true. The amount of tuition was generally low, and much lower than full-time earnings can be expected to be. For college students, average tuition was around \$1,800, while the average earnings of working 18-year-olds was \$ 8,600, and of working 22-year-olds was \$ 17,200. Therefore, the IRR we present is a lower bound to what it might be if individuals smoothed out their tuition expenditures.

Figure B-1: Propensity Score for Example of Doctorate vs Bachelor's degree, Males



Notes: The Propensity Score was estimated by probit, as part of the local linear regression matching procedure, using the standard covariates. The kernel density shown is just a smoothed representation of the histogram, with an Epanechnikov kernel and bandwidth .1.

Figure B-2: Three Estimation Methods,
for the Treatment Effect of Doctorate versus Bachelor's degree on Male Earnings



Notes: The causal effect of education on Terman males' earnings is determined by three methods: The nonparametric method uses local linear matching on the propensity score. The kernel used is tricube, bandwidth is .2 and the sample trims 1% of the observations. The corresponding distribution of the propensity score is shown in the next figure. The second model, parametric, allows coefficients on covariates to differ by treatment status. The common coefficient model includes only an indicator of treatment status, not an interaction. The covariates used for all methods are IQ, factor scores of Conscientiousness, Openness, and Extraversion at study start, education level of parents, their occupation, age at birth of subjects, parents' region of origin, the number of subjects' siblings, their birth order, their birth weight, whether they were breast fed, their current (1922) soundness of sleep, frequency of headaches, indicator of private tutoring, whether they participated/were involved in combat in WWII, and two indicators for cohort. We limit the sample to whites born from 1904-1915, who have at most 10 missing wage observations in the wage histories we constructed. The earnings are annual earnings in 2008 US Dollars, minus full tuition at the time it accrued, and after tax (rates for singles and married persons separately). See Appendix A for information on building the profiles from the raw data. The bootstrapped confidence interval is a basic 90% bootstrap confidence interval from the common coefficient model for 200 replications.