

BLS WORKING PAPERS

U.S. Department of Labor
U.S. Bureau of Labor Statistics
Office of Prices and Living Conditions



Ability Composition Effects on the Education Premium

Gregory Kurtzon, U.S. Bureau of Labor Statistics

Working Paper 456
July 2013

All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics.

Ability Composition Effects on the Education Premium

By Gregory Kurtzon*

Abstract

If unobserved ability is a significant portion of the college education premium, then a significant portion of the observed complementarity between the college and non-college educated is due to changes in the ability composition of those groups, overestimating the elasticity of complementarity up to 20%. If college attainment rose to over 50%, this effect would reverse, as is illustrated with high school attainment rates. If there is little ability bias, the distribution education related ability is nearly degenerate, with the awkward implication that the most productive individuals would earn barely more on average without a college education than the least.

1 Introduction

Workers with and without a college education are measured to be complements with each other in many studies including Katz and Murphy (1992), Bound and

*I would like to thank Randal Verbrugge, Ted To, Tim Erickson, Mike Lettau, Ralph Bradley, Paul Sullivan, Mina Kim, Judy Temple, Peter Meyer, Keenan Dworak-Fisher, Tricia Gladden, Rob McClelland, Ann Polivka, Thesia Garner, and Joshua Pinkston for useful comments and suggestions. All views expressed in this paper are those of the author and do not necessarily reflect the views or policies of the U.S. Bureau of Labor Statistics. Kurtzon: Bureau of Labor Statistics, 2 Massachusetts Ave., NE., Washington DC 20212 (e-mail: kurtzon.gregory@bls.gov), 202-691-6574.

Johnson (1993), and others discussed in Freeman (1986). This means that as the relative supply of highly educated workers rises, the relative demand for low education workers rises, and the relative wage of high vs. low education workers – the education premium – falls, and vice versa. The change in wages is generally considered to be due to a fall in the market price of the education level which had rising supply.

However, as surveyed in Card (1999), there is also a question among economists as to what extent education actually increases an individual's productivity, or is simply an indicator of greater inherent earning ability. Regressions of wages on schooling have a very significant coefficient on schooling. But the literature notes that to the extent that individuals with more unobserved earning ability are more likely to attain education, this spurious positive correlation between the error and the schooling variable, called "ability bias", would overestimate the causal effect of schooling on wages.

Consider if there was a significant amount of ability bias. If there were very few college graduates, they would on average be on the extreme right tail of the distribution of unobserved ability. Their relative wages and thus the measured education premium would be very high. If more individuals obtained a college education, the marginal individuals would on average be below the average ability of college graduates and above the average ability of non-college graduates. Therefore, this would lower the mean wages for both groups.¹ But because the mean high wage was already so high, it would fall proportionately more, and this effect of changing the ability composition of education groups would cause the college premium to fall just as if college and non-college graduates were complementary in production.

Interestingly, the effect would be reversed if at some time in the future most workers had a college education. Because there would be fewer non-college

graduates, the marginal workers obtaining education would reduce the non-college mean wages by proportionately more than the college mean wage, and the increase in relative supply would actually cause the college premium to rise. The rise in the share of the population graduating from high school to a majority over the later 20th Century is used to illustrate this.

Because composition changes act like changes in relative prices, measured elasticities of substitution would be biased. The effects of any market intervention that affected the relative quantity of educated labor with the goal of affecting wages would depend on these elasticities and be overestimated.

This model also offers an estimate of the distribution of unobserved ability given the degree of ability bias. This is used to demonstrate a very unrealistic conclusion of the assumption that there is little ability bias: that those who were the most likely to go to college would have earned almost the same on average without a college education as those who were the very least likely.

Section 2 develops the model and shows that under weak assumptions ability bias implies that ability composition effects would cause the premium to fall if less than half of the population was educated, but would cause the premium to rise if more than half was educated. Section 3 estimates the model with historical U.S. Census data and shows that if ability bias is significant, the composition effects would have been significant. It is also shown that if the highly educated are defined as high school graduates, as they were in the early 20th century, then the larger proportions of highly educated workers has led the composition effect to cause the high school premium to rise. The effects of low ability bias on the implied distribution of education related ability and the relative wages of the highest vs. lowest ability individuals without education is shown. Section 4 estimates the degree to which ignoring the composition effect would bias estimates of the complementarity between high and low education

workers. Section 5 concludes.

2 Model

As is typical in the literature, it is assumed that ability is positively correlated with education to some degree, i.e. that the marginal student to obtain a certain level of education has lower ability on average than others with that education and higher ability on average than those without it. One reason for this could be that the returns to education rise with ability.² Other reasons could be that higher ability students can more easily complete a college education, that they enjoy it more, or that they tend to have lower costs. Or, if colleges which subsidize education with public or private money don't expand enrollment to meet demand, as in Bound and Turner (2007), the college screening process selects the highest ability applicants to fill a number of slots for which there is excess demand.

Other studies have considered composition effects. Carneiro and Lee (2011) estimate a reduced form model of composition effects on the college premium, which is limited relative to a structural model.³ Juhn, Kim, and Vella (2005) present a structural model of schooling choice where it is likely that additional education will lower the average ability composition of the college educated relative to the population. But the model does not explicitly include the effects of ability sorting by education, and there is no effect of falling ability composition for non-college graduates. Therefore, they do not derive the pattern that composition effects would take, shown below.⁴

The wage model below is a very general framework that assumes little more than standard models. In any model of educational obtainment, there are a number of factors that will determine whether an individual is more or less likely to obtain a college education, such as the desire to attend college, the

individual cost, the expected return, interest rates, and others. Let X denote various individual-specific and market wide variables that effect the likelihood of an individual obtaining a college education, and let γ denote an individual's overall propensity to obtain a college education which is a function of these variables, so that

$$\gamma = f(X) \tag{1}$$

. Let the γ s for all individuals be indexed from highest to lowest by i , and let individuals be indexed by $i \in (0, 1)$, so that if N denotes the share of the population that obtains a college education, if $i \leq N$ then individual i obtains college.⁵

Consider a regression of log wages on schooling, similar to a stripped down version of the basic Mincer (1974) model, where w_i denotes the log wages of person i , S_i is a dummy variable denoting whether person i has a college education, b is the return to obtaining a college education, and x_i is the residual,

$$w_i = bS_i + x_i . \tag{2}$$

Therefore, x_i is earnings or earning ability not directly caused by a college education. It is reasonable to assume that x_i includes various factors including inherent earning ability and that this ability is correlated with γ , so that x_i is imperfectly correlated with S_i . In order to break out the component of x_i that is associated with schooling, the index i can be used. Let a_i denote the predicted value of x_i given i ,

$$a_i \equiv \hat{x}_i|i . \tag{3}$$

Since log wages w_i are approximately distributed as log-normal, it is reasonable

to use a normal distribution to predict x_i , so that

$$a_i \sim N(\mu_a, \sigma_a) , \quad (4)$$

where μ_a denotes the mean of both x_i and a_i . Therefore, if

$$G = \Phi\left(\frac{a - \mu_a}{\sigma_a}\right) \quad (5)$$

is the c.d.f. of a_i , then

$$a_i = G^{-1}(i) . \quad (6)$$

Let ε_i denote the prediction error:

$$\varepsilon_i \equiv x_i - a_i , \quad (7)$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon) ,$$

so that

$$x_i = a_i + \varepsilon_i .$$

Then equation (2) can be rewritten as

$$w_i = a_i + bS_i + \varepsilon_i . \quad (8)$$

This is a modification of a model from Card (1999), which in turn is a modification of Mincer (1974), where a_i represents unobserved ability. Ability bias in an estimation of (8) is when the omitted variable a_i is correlated with S_i , leading to bias in an OLS estimate of b . In this model, a_i represents the additional earnings that are correlated with the propensity to obtain college, even if person i doesn't do so, so one interpretation of a_i is that it's the expected

return to college attendance correlated ability.⁶

Since the top N percent of the population attends college, the top N percent of the distribution of a_i attends college. By the definition of a_i , ε_i includes all wage determining factors that are unrelated to a college education, and so is orthogonal to a_i and S_i .

This model does not require that all higher ability individuals go to college before lower ability individuals, and does not mean that individuals are perfectly sorted by ability into education groups. Only the expected education correlated ability is perfectly sorted. A low ability individual could have a low i because of a high desire to attend or a low cost of attendance. In this case, the ordering of x_i with respect to i would not be monotonic. To see this, consider an example of a simple economy where i is discrete, $i \in \{1, 2, 3, 4\}$.

i	x_i	a_i	ε_i
1	6.0344	6.1883	-0.1539
2	6.0653	5.9910	0.0743
3	5.9054	5.9505	-0.0451
4	5.964	5.7405	0.2235

Individual 2 has higher earning ability x_i than individual 1, even though 1 has a higher propensity to go to college and thus a higher *expected college related* earning ability, a_i . This could be because 1 had more desire or resources than 2, or because 2 has greater talents that aren't correlated with education (such as entertaining or athletic talents). But if averaged over individuals over a range of i , ability will tend to fall with a lower order in going to college. If the order in going to college has little relation to wages, then a large drop in i will on be associated with a small average drop in x_i , and thus a small drop in a_i . Therefore, the variance of a_i , σ_a , will be low. This allows for the possibility of a loose correlation between college related ability and average earnings. The

strength of this correlation is an empirical issue, and σ_a can be solved for with data for a given level of b , as shown in Section 3 below.

2.1 Composition Effects

Let distributions be approximated with a continuum of individuals and a_0 denote a_i for the cutoff individual for a college education, $a_0 \equiv G^{-1}(N)$. From (8) and the fact that ε_i has mean 0 and is uncorrelated with a_i , let

$$W^H \equiv \frac{\int_{a_0}^{\infty} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{a_0}^{\infty} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da} + b \quad (9)$$

$$W^L \equiv \frac{\int_{-\infty}^{a_0} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{-\infty}^{a_0} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da} \quad (10)$$

denote the mean college (H for high education) and non-college (L for low education) log wages in the population. Let the education premium P be defined as in the standard model as

$$P \equiv W^H - W^L = b^{OLS}, \quad (11)$$

where b^{OLS} is the OLS estimate of b in (2). The ability bias of this OLS estimation of equation (2) is simply the measured premium without the direct

effects of education, b , or

$$bias^{OLS} = b^{OLS} - b = \frac{\int_{a_0}^{\infty} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{a_0}^{\infty} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da} - \frac{\int_{-\infty}^{a_0} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{-\infty}^{a_0} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}, \quad (12)$$

which is the mean high education ability less the mean low education ability. This is derived in Appendix C. The greater the spread in mean abilities, such as from a higher σ_a , the greater the bias.

Consider the two extremes of college attainment: where almost no one is a college graduate, and where everyone is. In the first case, if N is small enough, the mean a_i of a college grad is arbitrarily large, so that $W^H = \infty$, while the mean non-college a_i approaches $\mu_a = W^L$. In the second case, W^H instead becomes $\mu_a + b$, while $W^L = -\infty$. Thus at both extremes, $P = \infty$.

Now consider what happens in the first case if N increased. The individuals with the highest a_i who did not attend college previously, around a_0 , would now attend. This would lower the mean a_i for both college graduates and non-college graduates. Since W^H was already arbitrarily large, it would tend to fall faster than W^L , pushing down the premium. This is illustrated in Figure 1.

This is exactly what would happen if college and non-college grads are complements – a relative increase in the quantity of one reduces its relative price. Because the composition effect mirrors the complementarity effect, studies such as those in Freeman (1986) and Katz and Murphy (1992) and Bound and Johnson (1992), which estimate the elasticity of substitution between college and non-college grads (or any division of workers into high and low educated) would *overestimate* the degree of complementarity.

But the two effects diverge when N grows larger than .5. When that happens, as W^L approaches $-\infty$ faster than W^H approaches μ_a , and the premium rises

again.

These effects are formalized in the following Theorem.

Theorem 1 *Suppose $a_i \sim f(a)$ is a two tailed non-degenerate symmetric distribution. In other words $a_0(N)$ is the college ability cutoff as a function of N , $\lim_{N \rightarrow 0} a_0(N) = \infty$ and $\lim_{N \rightarrow 1} a_0(N) = -\infty$, the mean is the median, $\mu_a = a_0(.5)$, and $\sigma_a > 0$. Also suppose $b \geq 0$. If $P(N)$ denotes the education premium as a function of N , then (1)*

$$1.1 \quad \lim_{N \rightarrow 1} P(N) = \infty \quad (13)$$

$$1.2 \quad \lim_{N \rightarrow 0} P(N) = \infty \quad (14)$$

and (2)

$$\arg \min P(N) = .5 . \quad (15)$$

Proof. See Appendix A. ■

Therefore, as more people obtain a college education, the premium will fall until half the working population has a college education, and then will rise after that. Graphs of examples of $P(N)$ as N changes are provided in Figures 2 and 3 in section 3 using distributions derived from the model and data. However, it should be noted that the Theorem is more general than the model and does not assume that the distribution of a_i is unimodal as the model does. Therefore, a_i could fall at a nonuniform rate and the Theorem would still hold.

Other ways to model the effects of education on wages are often used, and many consider them to be more intuitive. However, as shown in Appendix B, if they are consistent with the data they have similar implications to the above model, so the results are qualitatively robust to different specifications.

2.1.1 Other Measures of the Premium

Because using the premium defined as the difference in log means has this complementarity mimicking effect, one might think there was a better measure of the returns to education than this. But other potential measures have similar composition problems. Consider using the difference in median log wages between college and non college grads. When there are near zero college grads, the median high wage is still arbitrarily high, while the median low wage is near the population median, and vice versa when almost everyone is a graduate. The premium would still fall and then rise. Or consider the difference in non-log mean or median wages. When N is close to 0, the mean/median college wage is still arbitrarily high and the non-college close to the population mean/median, and thus the difference is arbitrarily high. When N is close to 1, the high mean/median goes to the population mean/median, and the low mean/median goes to 0, so the difference converges to the population mean/median. This has the same composition effect when $N < .5$, and a different one when $N > .5$.

3 Estimation

How important could composition effects reasonably be? To determine this, the identifiable parameters of the model are estimated with historical U.S. Census data, and the composition effects are measured under different assumptions about the unidentifiable parameters. The model is then reestimated defining "high education" as high school completion, in order to illustrate how composition effects can make the education premium rise.

Some other studies have measured ability composition effects. Carneiro and Lee (2011) find that the composition effects of increased college attendance lower the average wages of college graduates. They do this by comparing the

wages of individuals from regions that have different rates of college attendance, assuming skill prices are the same for all individuals in a region. However, the fall in the average wages of high school graduates due to composition effects is not significant at the 5% level. This could be because the true effect is small relative to the noise of the data and variables used.⁷

Juhn, Kim, and Vella (2005) run regressions of the relative wage of college grads to the population on supply relative to population, by cohort, to determine the magnitude of the composition effect. Empirically, it is impossible to tease apart the composition effect due to increased supply from the direct effect on education returns from increased supply. They use a predicted share of cohort education as the measure of composition, and various indirect measures of supply to control for the effects on education returns, and find composition effects to be very small. But this could be because these measures imperfectly capture the differences between composition and relative supply effects, and are still highly colinear with relative supply.

Rosenbaum (2003) also suggests that the ability composition of education may cause changes to the measured return to education, and runs regressions of mean wages with mean education rank for each education level as a control for unobserved ability. These show that composition effects significantly lower mean wages. But there are no clear implications for what the relative composition effects are on the education premium of college vs. non-college graduates because education levels were not divided into college vs. non-college graduates, but into four different groups.

Unlike the reduced form estimates of other studies, this model is structurally estimated. This allows for direct estimates and predictions of composition effects without proxy variables. But the data itself cannot separately identify the direct effect of education on wages, b , from ability bias. Therefore the amount of ability

bias must be assumed exogenously.

To see how the assumed ability bias affects the magnitude of composition effects, remember that from equations (11) and (12), ability bias is the difference between the measured premium and b , which means that it's the difference between the mean college a_i and mean non-college a_i . The composition effect is the change in the ability bias due to changes in a_0 . The higher b is assumed, the lower the ability bias and the lower the composition effects - there is a smooth trade-off between the composition effect and b . Therefore, as an upper bound on composition effects, the model is estimated under an extreme assumption about ability bias, that there is no causal effect of education on wages and the entire education premium is ability bias, i.e. $b = 0$ and $b^{OLS} = P$. If one assumes that $b = xb^{OLS}$, $0 < x < 1$, so that ability bias is $100 \cdot (1-x)\%$ of the premium, then the composition effects are $100 \cdot (1-x)\%$ of what is estimated under this bound. As an example of this, another bound is estimated which is on the other side of the span of opinions on ability bias. As stated in Card (1999), this opinion is that ability bias is quite small and could be on the same order as the effect of measurement error, which is estimated at an upward bias of about 10% of the direct effect of education.⁸ For simplicity, it will be assumed that education is measured accurately in the census and that the true b would be approximately 90% of the estimated coefficient in a simple regression of log wages on education, i.e. $b = .9b^{OLS}$. Finally, as an example of how the composition effect changes as the assumed ability bias changes, a middle point of 50% ability bias is also estimated, where $b = .5b^{OLS}$.

Census data from 1940 to 2000 was used.⁹ Only men who worked full time, 35 hours a week or more, were used. Also, to avoid issues involving individuals who are currently in school or different retirement ages for different educational groups, only individuals aged 25-65 were included. Weekly wages are used,

measured as annual labor income divided by annual weeks worked.

Parameters for each cohort were estimated separately for each census year. Since individuals usually get education when young, and are often competing against others in their cohort for college admissions (such as by high school class rank), the sorting by ability for college slots for a given year would mainly apply to the 18 year-old cohort for that year. The effect on the education premium of changes in the education levels of younger cohorts would be different than if we were to assume that changes in education levels were uniform across cohorts. For example, if older cohorts had little education and younger ones much more, overall the high mean wage would include high ability older workers and lower ability younger ones, and vice versa for the mean low wage. Estimating cohorts separately also allows for the measurement of purely demographic changes. Since younger workers tend to have a lower premium than older workers, changes in the average age of the workforce can change the overall premium.

The spread between the mean high wage and the mean overall wage, combined with the percent educated and the assumed ability bias, identifies the parameters of the college-related ability distribution and the ability cutoff for college. For a given N , the higher the relative mean high wage, the higher the variance in ability, and for a given relative mean high wage, the lower N is, the higher the variance in ability.

For each cohort c in each census year t , a regression of equation (2) was run, yielding an estimate of b of \hat{b}_{ct}^{OLS} . Then given the bound being used, $x = 0$ or $x = .9$, \hat{b}_{ct} is derived as

$$\hat{b}_{ct} = x\hat{b}_{ct}^{OLS} \tag{16}$$

. Then an estimate of the mean education correlated ability, μ_{act} , can be obtained from the unconditional mean wage \bar{w}_{ct} , \hat{b}_{ct} , and percent educated N_{ct}

by

$$\hat{\mu}_{act} = \bar{w}_{ct} - \hat{b}_{ct} N_{ct} . \quad (17)$$

Given the mean high education wage from the data W_{ct}^H , the standard deviation of ability and cutoff ability for each cohort c in each year t , $\hat{\sigma}_{act}$ and \hat{a}_{0ct} respectively, can be solved for numerically using

$$N_{ct} = \int_{\hat{a}_{0ct}}^{\infty} \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da \quad (18)$$

and

$$W_{ct}^H = \hat{b}_{ct} + \frac{\int_{\hat{a}_{0ct}}^{\infty} a \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da}{\int_{\hat{a}_{0ct}}^{\infty} \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da} . \quad (19)$$

Since a_i can be considered the return for college related ability for individual i , and b_{ct} is the direct return to education, $\hat{\sigma}_{act}$ and \hat{b}_{ct} can be considered to measure the returns for ability and education.

With these parameter estimates, equations (19) and

$$\hat{W}^H \left(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act} \right) = \hat{b}_{ct} + \frac{\int_{\hat{a}_{0ct}}^{\infty} a \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da}{\int_{\hat{a}_{0ct}}^{\infty} \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da} \quad (20)$$

$$\hat{W}^L \left(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act} \right) = \frac{\int_{\hat{a}_{0ct}}^{\hat{a}_{0ct}} a \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da}{\int_{-\infty}^{\hat{a}_{0ct}} \phi \left(\frac{a - \hat{\mu}_{act}}{\hat{\sigma}_{act}} \right) da} , \quad (21)$$

the model's predicted mean high and low education wages, $\hat{W}^H \left(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act} \right)$

and $\hat{W}^L(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act})$, can be generated as a baseline for comparison. Let θ_{ct}^E denote the percentage of individuals with education level $E \in \{H, L\}$ in cohort c in year t .¹⁰ Let a variable with a \sim and a \cdot subscript denote all values for that variable's subscript with the \cdot . So for example $\tilde{N}_{\cdot,t}$ denotes all values of N_{ct} for year t , while $\tilde{N}_{\cdot,\cdot}$ denotes all values of N_{ct} over all c and t . Then the overall premium in year t , P_t ,

$$\begin{aligned}
& P_t \left(\tilde{N}_{\cdot,t}, \tilde{b}_{\cdot,t}, \widetilde{\sigma}_{a,\cdot,t}, \widetilde{\mu}_{a,\cdot,t}, \tilde{\theta}_{\cdot,t}^H, \tilde{\theta}_{\cdot,t}^L \right) \\
&= \sum_c \theta_{ct}^H \hat{W}^H \left(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act} \right) - \sum_c \theta_{ct}^L \hat{W}^L \left(N_{ct}, \hat{b}_{ct}, \hat{\sigma}_{act}, \hat{\mu}_{act} \right) \quad (22)
\end{aligned}$$

is a function of both the distribution of ability and education's direct effect for each cohort, and also of the relative size of each cohort's educational group, θ_{ct}^E . This allows for the separate measurement of purely demographic effects on the premium as described in the next section. The model replicates the premium to four significant figures, which shows that the normality assumptions fit well.

3.1 Counterfactual Premia

The size of composition effects are measured by calculating what the premium would have been if it had only changed because of composition effects and comparing that to the actual premium. To do this, counterfactual premia for each census year, 1950-2000, were calculated by allowing the fraction N of educated individuals in each cohort to change, but not allowing the distribution of returns or the size of cohorts to change. To measure the total effect spanning the whole period, a counterfactual premium was also calculated for the fraction educated in 2000 with 1940's returns and demographic distributions. Also, counterfactual premia were calculated by allowing only demographics and then only the

returns to change to measure how large the effects of demographic and returns changes were.

Each set of counterfactual premia was calculated for the two bounds on b : $\hat{b}_{ct} = 0$ for all c and t , and $\hat{b}_{ct} = .9\hat{b}_{ct}^{OLS}$, and for $\hat{b}_{ct} = .5\hat{b}_{ct}^{OLS}$.

For $t \neq t'$, using $N_{ct'}$ for year t' and the ability distribution in year t to solve for a new counterfactual ability cutoff for year t , a_{0ct}^C , equations (18) and (19) yield

$$P_{tt'}^C \left(\tilde{N}_{.t'}, \tilde{b}_{.t}, \tilde{\sigma}_{a.t}, \tilde{\mu}_{a.t}, \tilde{\theta}_{.t}^H, \tilde{\theta}_{.t}^L \right) = \sum_c \theta_{ct}^H \hat{W}^H \left(N_{ct'}, \hat{b}_{ct}, \sigma_{act}, \hat{\mu}_{act} \right) - \sum_c \theta_{ct}^L \hat{W}^L \left(N_{ct'}, \hat{b}_{ct}, \sigma_{act}, \hat{\mu}_{act} \right) \quad (23)$$

the premium holding the demographic and returns distribution constant (to be referred to as the composition premium) between t and t' .

The premium measuring the effects of demographic change (to be referred to as the demographic premium) is

$$P_{tt'}^D \left(\tilde{N}_{.t}, \tilde{b}_{.t}, \tilde{\sigma}_{a.t}, \tilde{\mu}_{a.t}, \tilde{\theta}_{.t'}^H, \tilde{\theta}_{.t'}^L \right) = \sum_c \theta_{ct'}^H \hat{W}^H \left(N_{ct}, \hat{b}_{ct}, \sigma_{act}, \hat{\mu}_{act} \right) - \sum_c \theta_{ct'}^L \hat{W}^L \left(N_{ct}, \hat{b}_{ct}, \sigma_{act}, \hat{\mu}_{act} \right) \quad (24)$$

The effects of the distribution of returns (to be referred to as the returns premium) are measured by

$$P_{tt'}^W \left(\tilde{N}_{.t}, \tilde{b}_{.t'}, \tilde{\sigma}_{a.t'}, \tilde{\mu}_{a.t'}, \tilde{\theta}_{.t}^H, \tilde{\theta}_{.t}^L \right) = \sum_c \theta_{ct}^H \hat{W}^H \left(N_{ct}, \hat{b}_{ct'}, \sigma_{act'}, \hat{\mu}_{act'} \right) - \sum_c \theta_{ct}^L \hat{W}^L \left(N_{ct}, \hat{b}_{ct'}, \sigma_{act'}, \hat{\mu}_{act'} \right) . \quad (25)$$

The effects of each counterfactual change for a given decade can be seen by comparing the counterfactual premium for a given census year to the actual premium of the previous year. The effects over the whole period can be seen by comparing the 1940-2000 counterfactual premium to the actual 1940 premium.

Since a higher assumed b implies a proportionately lower composition effect, it also tends to imply a higher returns effect. To see this, note that if a large change in the actual premium cannot be explained with a large composition effect, and because the demographic premium uses period t mean wages which aren't affected by changes in composition or returns, it must be explained with a large change in returns. But since each change interacts with the others, there is no decomposition which can neatly divide the total change into the sum of the changes in the counterfactuals.

3.2 Results

3.2.1 Composition Effects as a Function of Percent Educated

Figure 2 is an example of how the premium varies due to the composition effect. The premium for 25 year old college grads in 1940 for $\hat{b}_{1915,1940} = 0$ is plotted against the percent educated, $N_{1915,1940}$, as an example. The premium can be seen to fall quickly away from 0 and 100, is symmetric around the median, and would fall by about 40% between the 1st/99th percentiles and the median due to the composition effect. This is roughly the mean decline of 41% across all cohorts and years, though cohort-years with a greater $\hat{\sigma}_{act}$ have steeper curves and larger differences between the 1st/99th percentiles and the median. The same mean percentage decline of about 41% holds for high school vs. non-high school grads, and the graphs have a similar shape.

When $\hat{b}_{1915,1940} = .9\hat{b}_{1915,1940}^{OLS}$, the curves are flatter and the total effects smaller because each $\hat{\sigma}_{act}$ is much smaller, as will be shown. The mean difference

between the 1st/99th percentile premium and the median is about 6% for both college and high school definitions. As shown in Figure 3 for the example of 25 year olds in 1940, the shape of the curve is similar.

3.2.2 College as Highly Educated

Figure 4 graphs N_{ct} . The percent college educated tends to rise across cohorts over time, but just as also documented in Card and Lemieux (2001), it fell for the peak of the baby boom. This can be seen in Figure 4 by the sharp declines in the latter part of the graphs for 1980, 1990, and 2000.

Tables 1-3 list the education premia for the 7 Census years. Over the entire period spanned, 1940-2000, the college premium rose only slightly, by 1.7%. As also documented in Goldin (1999), there was a very large fall in the premium in the 40s, but it returned over time and increased in every decade except the 70s.

First consider the case where $\hat{b}_{ct} = 0$. Figures 5 and 6 report the values of $\hat{\sigma}_{act}$ for 1940-1970 and 1980-2000 respectively. In this case, from equation (17), $\hat{\mu}_{act} = \bar{w}_{ct}$, and the values of \bar{w}_{ct} are graphed in Figures 7 and 8. The values of $\hat{\sigma}_{act}$ generally fall steeply in the 40s, and then rise every decade except the 70s. They also tend to rise in every year with age. This is consistent with the premium rising with age, and with the demographic effect of an aging population increasing the premium.

From Table 1, over the whole period the composition effect would have caused the premium to fall by about 20%. This is the cumulative effect of the declines due to composition for each decade. This is to be expected since N_{ct} is never above .5 and the overall percent educated rises every decade from 6.9% in 1940 to 29.8% in 2000, which is still on the steep part of the curve shown as an example in Figure 2. The composition effect was very small for the last two decades, since the percent college educated did not rise by much. The

effect of demographics would have caused the premium to rise by 10% over the whole period, which results from an increase each decade except the 70s, when large numbers of young people entered the workforce and moved the premium down. The effects of changing returns would have increased the premium by 14%. This combined effect was from an increase due to rising returns for each decade except the 40s and 70s. Thus, the combined effects of demographics and rising returns just barely edged out the downward composition effect.

Next consider the case where $\hat{b}_{ct} = .5\hat{b}_{ct}^{OLS}$, the intermediate between the two bounds. The values of $\hat{\sigma}_{act}$ under this assumption are exactly 50% of what they were when $\hat{b}_{ct} = 0$. Figures 9-12 graph the values of \hat{b}_{ct} and $\hat{\mu}_{act}$ under this assumption. From Table 2, the composition effect in this case is half as much as before, and would have caused a decline of 10% over the entire period, instead of 20%. The demographic effect is of course the same as before, since the breakdown of the correlation between wages and education by \hat{b}_{ct} versus $\hat{\sigma}_a$ does not affect it. The effects of changing returns therefore soak up the difference in the composition effects. Decade by decade they are qualitatively similar to before, but the increases are smaller than for $\hat{b}_{ct} = 0$, and would only have raised the premium slightly over the whole period, by about 2.7%.

Now consider the bound where $\hat{b}_{ct} = .9\hat{b}_{ct}^{OLS}$, or when ability bias is only 10%. The values of $\hat{\sigma}_{act}$ under for this bound are now exactly 10% of what they were when $\hat{b}_{ct} = 0$. Figures 13-16 graph the values of \hat{b}_{ct} and $\hat{\mu}_{act}$ under this assumption. From Table 3, the composition effect in this case is very small, and would only have caused a decline of 2% over the entire period, instead of 20%. Decade by decade the effects of changing returns are also qualitatively similar to before, but the increases are smaller during the 50s through 70s when the composition effect was the largest for $\hat{b}_{ct} = 0$. So over the whole period the effects of different returns would actually have caused the premium to fall by

6%, as the decline in the 40s would not have been completely reversed by 2000.

The assumption of only 10% ability bias makes for a very extreme assumption about education related ability: that there is almost no inherent ability associated with education. For example, consider the cohort born in 1960. By 2000, if there was no college education available, those in the top 10% of the ability distribution who would have gone to college would only have earned 13.5% more on average than the bottom 10% of the distribution, which would be most of the high school dropouts. The top 5% of the ability distribution, who would have attended the most prestigious universities, would only have earned 17.7% more than the bottom 5%. Under this assumption, colleges go through an extensive selection process simply to find students who would have been barely more productive than the average individual.

3.2.3 High School as Highly Educated

Early in the 20th century, the status of a high school graduate was similar to the status of a college graduate today.¹¹ High school graduates can be used as the definition of highly educated as a way to study how the composition effect would act if college attainment rates rose above 50%. While high schools usually do not screen applicants, it is still a reasonable assumption that the least productive students are the ones who don't graduate, so that ability is still sorted into education.

As shown in Figure 17, the percent of the population by cohort with a high school education, N_{ct}^{HS} , rose significantly for every census year nearly uniformly, with younger cohorts having a higher graduation rate. In 1940, all cohorts were below 50%, but by 1980 they were all above. Shown in Tables 4-6, the premium rose by approximately 23% over the entire period, and followed a similar pattern to the college premium, falling in the 40s and 70s and rising every other decade.

For the bound of $\hat{b}_{ct} = 0$ shown in Figures 18 and 19, the values of $\hat{\sigma}_{act}$ for high school tend to follow the same pattern as the premium, falling in the 40s and 70s and rising at other times, but don't surpass the 1940 levels. They are also roughly of the same magnitude as the standard deviations of college related ability.

Table 4 shows the counterfactual premia for high school graduates as highly educated for this bound. In contrast to the composition effects of college education, and because the fraction of high school graduates is higher, the composition effect would have caused the high school premium to rise over the whole period by 18%. For the first two decades, the 40s and 50s, it would have caused the premium to fall, since most cohorts still had less than 50% graduating. But the later increases overwhelmed this initial fall. Demographics would have caused the premium to rise by 20% in all three assumptions for \hat{b}_{ct} , due to the population aging in every decade except for the entrance of the baby boom into the workforce in the 60s and 70s. Changing returns would have reduced the premium by 13%, following a similar qualitative pattern decade by decade to college returns except that the fall in the 40s was not made up for by later rises.

When ability bias is 50%, the standard deviations, graphed in Figures 20 and 21, are of course 50% of what they were before, just as with college. The values of \hat{b}_{ct} and $\hat{\mu}_{act}$ under this case are graphed in Figures 22-25, and Table 5 shows the counterfactual premia. The composition effect reversed in the 60s, as under the $\hat{b}_{ct} = 0$ bound, but the effects are all half as much. Over the whole period, composition effects would have made the premium rise by 9% instead of 18%, and the effects of changing returns would have been qualitatively similar and reduced the premium by only 4.6%.

When ability bias is assumed to be 10%, the standard deviations, graphed in Figures 26 and 27, are of course 10% of what they were under $\hat{b}_{ct} = 0$, just as

with college. The values of \hat{b}_{ct} and $\hat{\mu}_{act}$ under this assumption are graphed in Figures 28-31. Even though it wouldn't necessarily be the case that the ability bias for high school is the same as college, this is still a large assumption. For the example of the premium in 2000 of the cohort born in 1960, if there was no high school education, those who would have had a high school education would only have earned 5% more than those who wouldn't have. This is a very small difference considering that the dropouts were the bottom 14% of that cohort in 2000. The top 5% of high school graduating classes would only have earned 11.6% more than the bottom 5%, or approximately the bottom 1/3 of dropouts.

Table 6 shows the high school counterfactuals for 10% ability bias. All three effects over the whole period contributed to make the premium rise. The composition effect made the premium rise by less of course, by about 2%, about the same as from rising returns. Also as before, much more of the overall change in the premium is attributed to the change in returns since the composition effect is so much smaller.

4 The Overestimation of Complementarity between High and Low Education Workers

So what does this imply for the measured complementarity between high and low education workers?

The literature on this complementarity, measured by the elasticity of substitution between them, was surveyed by Freeman (1986). Cross sectional studies comparing different countries or states that showed very high or infinite elasticities were followed by time series studies that were more reliable and showed lower elasticities, around .4 - 3.5, usually around 1 to 2. Later, Katz and Murphy (1992) and Bound and Johnson (1992) estimated the elasticity by regressing the

education premium on the ratio of relative supplies using data from the Current Population Survey (CPS) in the U.S., using slightly different methods and obtaining estimates of 1.4 and 1.7 respectively.

If demand can be represented by a CES function, as is typically assumed in these studies, then

$$\log\left(\frac{W_t^H}{W_t^L}\right) = \frac{1}{\eta}\text{shift}_t - \frac{1}{\eta}\log\left(\frac{N_t^H}{N_t^L}\right), \quad (26)$$

where η is the elasticity of substitution between high and low education workers, W_t^H , W_t^L , N_t^H , and N_t^L are mean wages and total quantities for year t , and shift_t is the effect of all other factors affecting relative wages for year t . The shift term is important as it has been well documented in the aforementioned studies that relative wages cannot be explained by supply changes alone, and that there is a strong trend for rising relative college wages. The shift term has usually been assumed to denote an increasing relative demand for college graduates.

Equation (26) can be estimated in a similar method as other studies by taking the first differences and assuming the change in demand is constant,¹² with

$$\left[\log\left(\frac{W_t^H}{W_t^L}\right) - \log\left(\frac{W_{t-1}^H}{W_{t-1}^L}\right)\right] = \text{constant} - \frac{1}{\eta}\left[\log\left(\frac{N_t^H}{N_t^L}\right) - \log\left(\frac{N_{t-1}^H}{N_{t-1}^L}\right)\right] + v_t \quad (27)$$

where v_t is error.

In order to remove the composition effect, a new counterfactual is calculated which allows the returns and demographics to change, but not the percent of

each cohort college educated, $\tilde{N}_{.t}$,

$$\begin{aligned}
& P_{t,t'}^{DR} \left(\tilde{N}_{.t}, \tilde{b}_{.t'}, \tilde{\sigma}_{a.t'}, \tilde{\mu}_{a.t'}, \tilde{\theta}_{.t}^H, \tilde{\theta}_{.t}^L \right) \\
&= \sum_c \theta_{ct'}^H \hat{W}^H (N_{ct}, b_{ct'}, \hat{\sigma}_{act'}, \hat{\mu}_{act'}) \\
&\quad - \sum_c \theta_{ct'}^L \hat{W}^L (N_{ct}, b_{ct'}, \hat{\sigma}_{act'}, \hat{\mu}_{act'}) . \tag{28}
\end{aligned}$$

The first term is replaced by the change that the premium would have had if there was no composition effect,

$$\begin{aligned}
& \left[P_{t,t'}^{DR} \left(\tilde{N}_{.t}, \tilde{b}_{.t'}, \tilde{\sigma}_{a.t'}, \tilde{\mu}_{a.t'}, \tilde{\theta}_{.t}^H, \tilde{\theta}_{.t}^L \right) - \log \left(\frac{W_{t-1}^H}{W_{t-1}^L} \right) \right] \\
&= \text{constant} - \frac{1}{\eta} \left[\log \left(\frac{N_t^H}{N_t^L} \right) - \log \left(\frac{N_{t-1}^H}{N_{t-1}^L} \right) \right] + v_t . \tag{29}
\end{aligned}$$

To set a bound on how large this effect could be, the counterfactual is calculated under the assumption that there is no direct effect of education, i.e. $b = 0$. As in the above studies, the March CPS supplements from 1964-2009 is used because the Census does not have enough data points for a regression, and also to make the results more comparable.¹³ The mean wages W_t^E are the geometric means of wages in the CPS for education level E in year t , and N_t^E is the total number of individuals of education level E in year t .

The results are reported in Table 7 below.

Table 7: Regression Estimates of Substitution

Elasticity between High and Low Educated Workers

Controlling for Composition Effect (standard errors in parentheses)

Variable	Model			
	1	2	3	4
Constant	-60.488 (9.078)			
Trend	.030 (.005)	.0165 (.006)	.0129 (.004)	.014 (.004)
Difference in Relative Quantities	-.713 (.108)	-.309 (.105)	-.280 (.086)	-.231 (.088)

For comparison, model (1) is similar to Katz and Murphy (1992), where demand is represented as a linear time trend, and only uses CPS data before 1988 as they did. The results are nearly equal. Since controlling for composition effects requires using first differenced data, the effects of first differencing must be measured. Model (2) is similar to (1) except for using first differenced data for comparison, so that the constant becomes the trend term. This does make a significant difference, nearly cutting both coefficients in half. This means the implied elasticity of substitution is 3.2 instead of 1.4. Although this is still within the wide range of elasticity estimates, such a large difference is indicative of a poorly specified model. Model (3) is the same as (2) except that it uses all the available data, from 1964-2009, and the coefficients are nearly the same.

Model (4) is equation (29), and comparison with model (3) shows what happens when the composition effect is removed. The relative supply coefficient falls by about 20%, meaning that ignoring the composition effect could lead to an overestimation of the elasticity of up to 20% over this period. These results are robust to the inclusion of an autoregressive term, using second differences, and changing the period covered.

5 Conclusions/Discussion

This paper has shown that if ability bias is significant, then the ability composition effects of increased education are significant. A higher assumed ability bias implies a proportionately higher composition effect. The effect could have reduced the measured education premium by 2-20% over the late 20th century, and biased the measured elasticity between high and low education workers by up to 20%.¹⁴

However, college education levels have only risen slowly since the mid 1970s, and if these levels should rise more rapidly, the total composition effects would be more important. If those levels ever went above 50%, a testable implication of significant ability bias would be that further increases in education cause the premium to rise, just as the rise in high school education caused the high school premium to rise. Even in this unlikely event, it would still be difficult to tease apart any demand shifts, since without those, the premium has already risen along with increases in the supply of college educated workers.

The composition effect would be important in studying the effects of promoting additional education. It could also be an important component in signaling models of education, which require composition shifts to explain why the value of the signal changes with relative supply.

It has also been illustrated that the assumption of little ability bias implies a nearly degenerate distribution of college correlated ability, so that the very top college graduates would have earned almost the same on average without college as the very lowest ability workers. That implies that the propensity to go to college has almost nothing to do with earning ability, which is hard to reconcile with the intense college application and screening process.

6 Tables

Table 1: Counterfactual College Education Premia Assuming no

Census Base Year	Direct Causal Effect of Education			
	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.4413	0.6037	0.6223
1940	0.5465			
1950	0.3507	0.5289	0.5500	0.3543
1960	0.4099	0.3340	0.3518	0.4281
1970	0.4565	0.3902	0.4240	0.4672
1980	0.3576	0.4317	0.4470	0.3890
1990	0.4873	0.3530	0.4028	0.4506
2000	0.5558	0.4858	0.4965	0.5550

Table 2: Counterfactual College Education Premia Assuming 50%

Census Base Year	Ability Bias			
	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.4939	0.6037	0.5614
1940	0.5465			
1950	0.3507	0.5377	0.5500	0.3487
1960	0.4099	0.3424	0.3518	0.4180
1970	0.4565	0.4000	0.4240	0.4559
1980	0.3576	0.4441	0.4470	0.3781
1990	0.4873	0.3553	0.4028	0.4486
2000	0.5558	0.4865	0.4965	0.5544

Table 3: Counterfactual College Education Premia Assuming 10%

Census Base Year	Ability Bias			
	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.5360	0.6037	0.5127
1940	0.5465			
1950	0.3507	0.5447	0.5500	0.3440
1960	0.4099	0.3489	0.3518	0.4100
1970	0.4565	0.4079	0.4240	0.4469
1980	0.3576	0.4540	0.4470	0.3694
1990	0.4873	0.3571	0.4028	0.4469
2000	0.5558	0.4872	0.4965	0.5540

Table 4: Counterfactual High School Education Premia Assuming

Census Base Year	no Direct Causal Effect of Education			
	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.4836	0.4921	0.3546
1940	0.4088			
1950	0.2811	0.3947	0.4153	0.2840
1960	0.3165	0.2785	0.2877	0.3093
1970	0.3432	0.3231	0.3144	0.3381
1980	0.3113	0.3660	0.3371	0.3016
1990	0.4279	0.3401	0.3591	0.3408
2000	0.5012	0.4470	0.4742	0.4427

Table 5: Counterfactual High School Education Premia Assuming

50% Ability Bias				
Census Base Year	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.4462	0.4921	0.3900
1940	0.4088			
1950	0.2811	0.4018	0.4153	0.2794
1960	0.3165	0.2798	0.2877	0.3078
1970	0.3432	0.3198	0.3144	0.3418
1980	0.3113	0.3546	0.3371	0.3125
1990	0.4279	0.3257	0.3591	0.3575
2000	0.5012	0.4375	0.4742	0.4525

Table 6: Counterfactual High School Education Premia Assuming

10% Ability Bias				
Census Base Year	Actual Premium	Composition Change	Demographic Change	Returns Change
1940-2000		0.4163	0.4921	0.4184
1940	0.4088			
1950	0.2811	0.4074	0.4153	0.2757
1960	0.3165	0.2809	0.2877	0.3066
1970	0.3432	0.3171	0.3144	0.3447
1980	0.3113	0.3455	0.3371	0.3270
1990	0.4279	0.3220	0.3591	0.3709
2000	0.5012	0.4298	0.4742	0.4604

7 Appendix A. :Proof of Theorem

Proof. 1.1

$$\lim_{N \rightarrow 1} W^H = \lim_{N \rightarrow 1} \frac{\int_{a_0(N)}^{\infty} af(a) da}{\int_{a_0(N)}^{\infty} f(a) da} + b = \lim_{a_0 \rightarrow -\infty} \frac{\int_{a_0}^{\infty} af(a) da}{\int_{a_0}^{\infty} f(a) da} + b = \frac{\int_{-\infty}^{\infty} af(a) da}{\int_{-\infty}^{\infty} f(a) da} + b = \mu_a + b ,$$

and

$$\lim_{N \rightarrow 1} W^L = \lim_{N \rightarrow 1} \frac{\int_{-\infty}^{a_0(N)} af(a) da}{\int_{-\infty}^{a_0(N)} f(a) da} = \lim_{N \rightarrow 1} \frac{\int_{-\infty}^{a_0(N)} af(a) da}{\int_{-\infty}^{a_0(N)} f(a) da} = \lim_{a_0 \rightarrow -\infty} \frac{\int_{-\infty}^{a_0} af(a) da}{\int_{-\infty}^{a_0} f(a) da} .$$

By L'Hospital's rule,

$$\lim_{a_0 \rightarrow -\infty} \frac{\int_{-\infty}^{a_0} af(a) da}{\int_{-\infty}^{a_0} f(a) da} = \lim_{a_0 \rightarrow -\infty} \frac{a_0 f(a_0)}{f(a_0)} = -\infty .$$

Therefore $\lim_{N \rightarrow 1} P(N) = \lim_{N \rightarrow 1} (W^H - W^L) = \mu_a + b + \infty = \infty$.

1.2

$$\lim_{N \rightarrow 0} W^H = \lim_{N \rightarrow 0} \frac{\int_{a_0(N)}^{\infty} af(a) da}{\int_{a_0(N)}^{\infty} f(a) da} = \lim_{a_0 \rightarrow \infty} \frac{\int_{a_0}^{\infty} af(a) da}{\int_{a_0}^{\infty} f(a) da} .$$

By L'Hospital's rule,

$$\lim_{a_0 \rightarrow \infty} \frac{\int_{a_0}^{\infty} a f(a) da}{\int_{a_0}^{\infty} f(a) da} = \lim_{a_0 \rightarrow \infty} \frac{-a_0 f(a_0)}{-f(a_0)} = \infty .$$

and

$$\lim_{N \rightarrow 0} W^L = \lim_{N \rightarrow 0} \frac{\int_{-\infty}^{a_0(N)} a f(a) da}{\int_{-\infty}^{a_0(N)} f(a) da} = \lim_{a_0 \rightarrow \infty} \frac{\int_{-\infty}^{a_0} a f(a) da}{\int_{-\infty}^{a_0} f(a) da} = \frac{\int_{-\infty}^{\infty} a f(a) da}{\int_{-\infty}^{\infty} f(a) da} = \mu_a .$$

Therefore $\lim_{N \rightarrow 0} P(N) = \lim_{N \rightarrow 0} (W^H - W^L) = \infty - \mu_a = \infty$.

2. Differentiating the premium w.r.t. a_0 yields the F.O.C.

$$\frac{\partial}{\partial a_0} (W_H - W_L) = 0 \tag{30}$$

→

$$f(a_0) \left[\frac{\int_{a_0}^{\infty} a f(a) da}{\left(\int_{a_0}^{\infty} f(a) da \right)^2} + \frac{\int_{-\infty}^{a_0} a f(a) da}{\left(\int_{-\infty}^{a_0} f(a) da \right)^2} - \frac{a_0}{\int_{a_0}^{\infty} f(a) da} - \frac{a_0}{\int_{-\infty}^{a_0} f(a) da} \right] = 0 \tag{31}$$

→

$$(1 - N) W_H + N W_L - a_0 = 0 , \tag{32}$$

which is solved at the point $N = .5$ if $a_0(.5) = \mu_a$. The S.O.C. are

$$\frac{\partial^2}{\partial a_0^2} (W_H - W_L) = 0$$

→

$$2f(a_0)^2 \left[\frac{\int_{a_0}^{\infty} af(a) da}{\left(\int_{a_0}^{\infty} f(a) da\right)^3} - \frac{\int_{-\infty}^{a_0} af(a) da}{\left(\int_{-\infty}^{a_0} f(a) da\right)^3} + \frac{a_0}{\left(\int_{-\infty}^{a_0} f(a) da\right)^2} - \frac{a_0}{\left(\int_{a_0}^{\infty} f(a) da\right)^2} \right] = 0 \quad (33)$$

→

$$\frac{2f(a_0)^2}{N^2(1-N)^2} \left[(W^H - a_0)(1-N)^2 + (a_0 - W^L)N^2 \right] > 0 \quad (34)$$

for $0 < N < 1$, $\sigma_a > 0$, since $W^H - a_0 > 0$ and $a_0 - W^L > 0$ if $\sigma_a > 0$. Therefore $N = .5$ is a unique minimum. ■

8 Appendix B. : Alternative Specifications

8.1 Heterogeneous Returns to Education

Another modification of the model assumes that the return to education differs among individuals,

$$w_i = a + b_i S_i + \varepsilon_i \quad (35)$$

$$b_i \sim N(\mu_b, \sigma_b) \quad (36)$$

and $b_i \perp \varepsilon_i$. In this model, those individuals with the highest return to education would have the highest propensity to obtain it: the highest N percent of the

distribution of b_i would obtain a college education. Therefore,

$$W^H = a + \frac{\int_{b_0}^{\infty} b \phi\left(\frac{b-\mu_b}{\sigma_b}\right) db}{\int_{b_0}^{\infty} \phi\left(\frac{b-\mu_b}{\sigma_b}\right) db} \quad (37)$$

$$W^L = a, \quad (38)$$

and

$$P(N) = W^H - W^L = \frac{\int_{b_0}^{\infty} b \phi\left(\frac{b-\mu_b}{\sigma_b}\right) db}{\int_{b_0}^{\infty} \phi\left(\frac{b-\mu_b}{\sigma_b}\right) db}, \quad (39)$$

meaning that the premium is simply the mean return to college of the top N percent of the distribution. By the same arguments in the Theorem about the mean high wage, W^H , the premium will decline from ∞ to μ_b as $N_{0 \rightarrow 1}$. This is similar to using differences in mean or median wages – there is still a composition effect mimicking complementarity, but there is no reverse effect causing the premium to rise again. This model is not consistent with the data, however, since it implies the distribution of non-college wages is degenerate.

8.2 Heterogeneous Ability and Returns to Education

Now consider a combination of the two models, which is a simplification of a model in Card (1999) without the quadratic term in schooling,

$$w_i = a_i + b_i S_i + \varepsilon_i \quad (40)$$

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim N \left(\begin{bmatrix} \mu_a \\ \mu_b \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} \right), \quad (41)$$

$a_i, b_i \perp \varepsilon_i$, with the propensity to go to college assumed to be correlated with the return to college, so that $\sigma_{ab} > 0$. It is still the case that the highest ability individuals obtain education. But while on average they have the highest return to education, it is not always true.

Now,

$$W^H = \frac{\int_{a_0}^{\infty} a f(a) da}{\int_{a_0}^{\infty} f(a) da} + \mu_b + \frac{\sigma_{ab}}{\sigma_a} \frac{\phi\left(\frac{a_0 - \mu_a}{\sigma_a}\right)}{1 - \Phi\left(\frac{a_0 - \mu_a}{\sigma_a}\right)} \quad (42)$$

$$W^L = \frac{\int_{-\infty}^{a_0} a f(a) da}{\int_{-\infty}^{a_0} f(a) da} \quad (43)$$

$$W^H - W^L = \frac{\int_{a_0}^{\infty} a f(a) da}{\int_{a_0}^{\infty} f(a) da} - \frac{\int_{-\infty}^{a_0} a f(a) da}{\int_{-\infty}^{a_0} f(a) da} + \mu_b + \frac{\sigma_{ab}}{\sigma_a} \frac{\phi\left(\frac{a_0 - \mu_a}{\sigma_a}\right)}{1 - \Phi\left(\frac{a_0 - \mu_a}{\sigma_a}\right)} \quad (44)$$

As before, $\lim_{N \rightarrow 1} W^H = \infty$, $\lim_{N \rightarrow 0} W^L = -\infty$, and the premium again moves down and then up from composition effects. Since a_i and b_i are not directly observed, and μ_a , μ_b , σ_a , and σ_{ab} cannot be separately identified from data on wages and education, this model is not empirically distinguishable from the original model. Since it has qualitatively similar implications the simpler model can be used without loss of generality.

9 Appendix C.: Bias of OLS in the Model

Consider breaking a_i into

$$a_i = \mu_a + \alpha_i \quad (45)$$

where α_i is the deviation from mean ability. Then (8) becomes

$$w_i = \mu_a + \alpha_i + bS_i + \varepsilon_i \quad (46)$$

Since the OLS bias from an omitted variable, X_2 , is $(X_1'X_1)^{-1}X_1'X_2\beta_2$, where X_1 denotes the included independent variables, β_2 is the coefficient on the omitted variable, which on α_i is 1 here, taking deviations from means (the mean of S_i being N) to drop the constant makes this

$$([S_i - N]'[S_i - N])^{-1}[S_i - N]'\alpha_i \quad (47)$$

Let N_H denote the total number of college graduates, and N_L denote the total number of non-college graduates, so that

$$N = \frac{N_H}{N_H + N_L} .$$

Then $([S_i - N]'[S_i - N])^{-1} = \frac{N_H N_L}{N_H + N_L}$, and $[S_i - N]'\alpha_i = \sum_i (S_i - N)\alpha_i$, which can be divided into separate sums for college and non-college grads, $\sum_{S_i=1} (1 - N)\alpha_i + \sum_{S_i=0} (-N)\alpha_i$. This becomes $\frac{N_H N_L}{N_H + N_L} \left(\frac{\sum_{S_i=1} \alpha_i}{N_H} - \frac{\sum_{S_i=0} \alpha_i}{N_L} \right)$, so that the bias in the sample is $\frac{\sum_{S_i=1} \alpha_i}{N_H} - \frac{\sum_{S_i=0} \alpha_i}{N_L}$, which for a continuum, if

$$a_0 = \mu_a + \alpha_0 , \quad (48)$$

is

$$\frac{\int_{\alpha_0+\mu_a}^{\infty} \alpha \phi\left(\frac{\alpha}{\sigma_a}\right) d\alpha}{\int_{\alpha_0+\mu_a}^{\infty} \phi\left(\frac{\alpha}{\sigma_a}\right) d\alpha} - \frac{\int_{-\infty}^{\alpha_0+\mu_a} \alpha \phi\left(\frac{\alpha}{\sigma_a}\right) d\alpha}{\int_{-\infty}^{\alpha_0+\mu_a} \phi\left(\frac{\alpha}{\sigma_a}\right) d\alpha} = \frac{\int_{a_0}^{\infty} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{a_0}^{\infty} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da} - \frac{\int_{-\infty}^{a_0} a \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da}{\int_{-\infty}^{a_0} \phi\left(\frac{a-\mu_a}{\sigma_a}\right) da} .$$

Notes

¹Evidence for this effect on the mean wages of college graduates was found by Carneiro & Lee (2011), discussed below.

²An alternative form of the model with this feature is shown in Appendix B.

³While they show that their results are qualitatively robust to the choice of function, because they use a reduced form they are not quantitatively robust for edge of sample and out of sample predictions. For example, the predicted effect on the education premium when the proportion with some college goes from 30% to 40% is roughly twice the size in one specification compared to another. They also have no prediction of the composition effect reversing.

⁴The model that they estimate is a reduced form that is not derived by the paper's economic model.

⁵It is assumed that the ordering i is roughly stable over time.

College is considered here to be a bachelor's degree or the equivalent.

⁶A high value of a_i does not necessarily represent intelligence or productivity. For example, a low value of i (and thus high a_i) could be due to parental wealth, which could be correlated with social networks that boost earnings.

⁷As shown above, when a certain level of educational attainment is high (40% to 70% in their data for those with no college), the mean wage is close to the population mean and the composition effects would be small. The data and variables are not perfect measures because: (i) there is little variation in attendance between regions; (ii) they use the mean wage for those with exactly 12 years of schooling instead of 12 years or less, as in this paper; (iii) they use the lifetime average proportion educated for each cohort instead of the proportion educated in a given year.

⁸Kane, Rouse, and Staiger (1999) have estimated that 90% of individuals with a bachelor's degree accurately report it.

⁹Data was obtained from the Integrated Public Use Microdata Sample (IPUMS) website.

¹⁰ θ_{ct}^E is the total number of education level E individuals in cohort c in year t divided by the total number of individuals of education level E in year t .

¹¹See Goldin (1998).

¹²Katz & Murphy (1992) don't use a first difference, while Bound & Johnson (1992) use second differences in a fixed effects model, with more than two types of workers being substituted. Katz & Murphy (1992) also use arithmetic means for the mean wages and instead of dividing workers into college and non-college educated, they allocated proportions of those

with some college to either college or high school graduates.

¹³Education data is not available in the March 1963 CPS, and because adjacent years are needed, 1962 and 1963 are not used.

¹⁴Since measurement error was not controlled for, it may be more accurate to say that the composition effect reduced the premium by 2.2-22%. This assumes that measurement error implies that the true premium is 10% higher, so σ_a and thus the composition effect are 10% higher.

10 References

1. Bound, John, and Johnson, George. "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations." *The American Economic Review* 82 (1992): 371-392.
2. Bound, John and Turner, Sarah. "Cohort crowding: How resources affect collegiate attainment." *Journal of Public Economics* 91 (2007): 877-899.
3. Cameron, Stephen and Taber, Christopher. "Borrowing Constraints and the Returns to Schooling." NBER Working Paper 7761 (2000).
4. Card, David. "The Causal Effect of Education on Earnings." in *The Handbook of Labor Economics* Vol. 3 (1999): 1801-1863.
5. Card, David, and Lemieux, Thomas. "Dropout and Enrollment Trends in the Post-War Period: What Went Wrong in the 1970s?" in *Risky Behavior among Youths: An Economic Analysis* (2001).
6. Carneiro, Pedro, and Lee, Sokbae. "Trends in Quality-Adjusted Skill Premia in the United States, 1960-2000." *American Economic Review* 101 (2011): 2309-2349.
7. Freeman, Richard B. "Demand for Education." in *The Handbook of Labor Economics* Vol. 1 (1986): 357-386.

8. Goldin, Claudia. "America's Graduation from High School: The Evolution and Spread of Secondary Schooling in the Twentieth Century." *The Journal of Economic History* 58 (1998): 345-374.
9. Goldin, Claudia. "A Brief History of Education in the United States." NBER Historical Paper 119 (1999).
10. Juhn, Chinhui, Dae IL Kim, and Francis Vella. "The Expansion of College Education in the United States: Is There Evidence of Declining Cohort Quality?" *Economic Inquiry* 43 (2005): 303-315.
11. Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101 (1993): 410-442.
12. Katz, Lawrence F. and Murphy, Kevin. "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *The Quarterly Journal of Economics* 107 (1992): 35-87.
13. Mincer, Jacob. *Schooling, Experience, and Earnings* (1974).
14. Murphy, Kevin M. and Finis Welch. "The Structure of Wages." *The Quarterly Journal of Economics* 107 (1992): 285-326.
15. Rosenbaum, Dan T. "Ability, Education Ranks, and Labor Market Trends: The Effects of Shifts in the Skill Composition of Educational Groups." Unpublished paper, 2003.

Figure 1: Distribution of a and Cutoff for Education, a_0

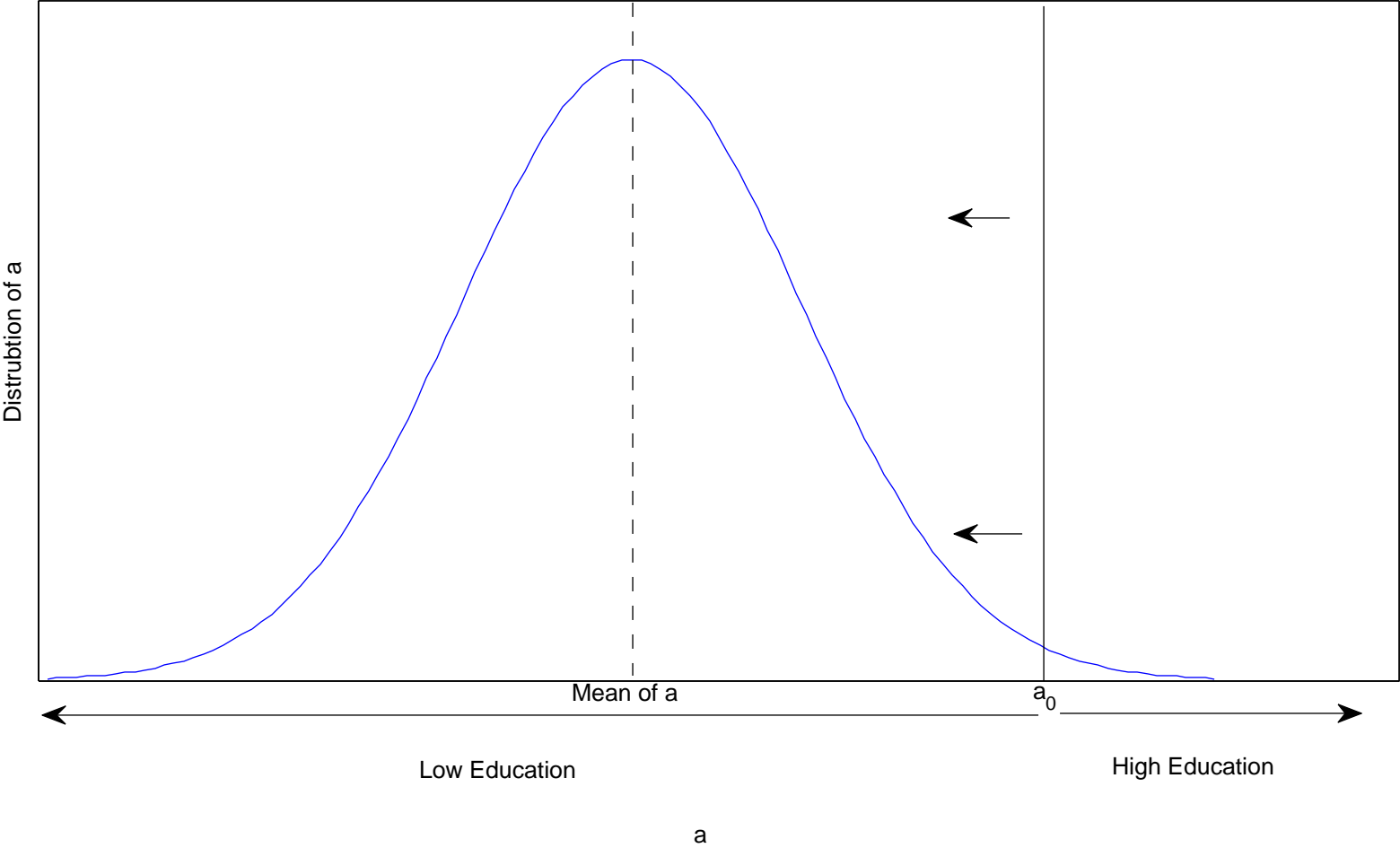


Figure 2: Counterfactual Education Premium as a Function of Percent of Population with a College Education due to Composition Effects, for 25 Year Olds in 1940, for $b_{1915,1940} = 0$

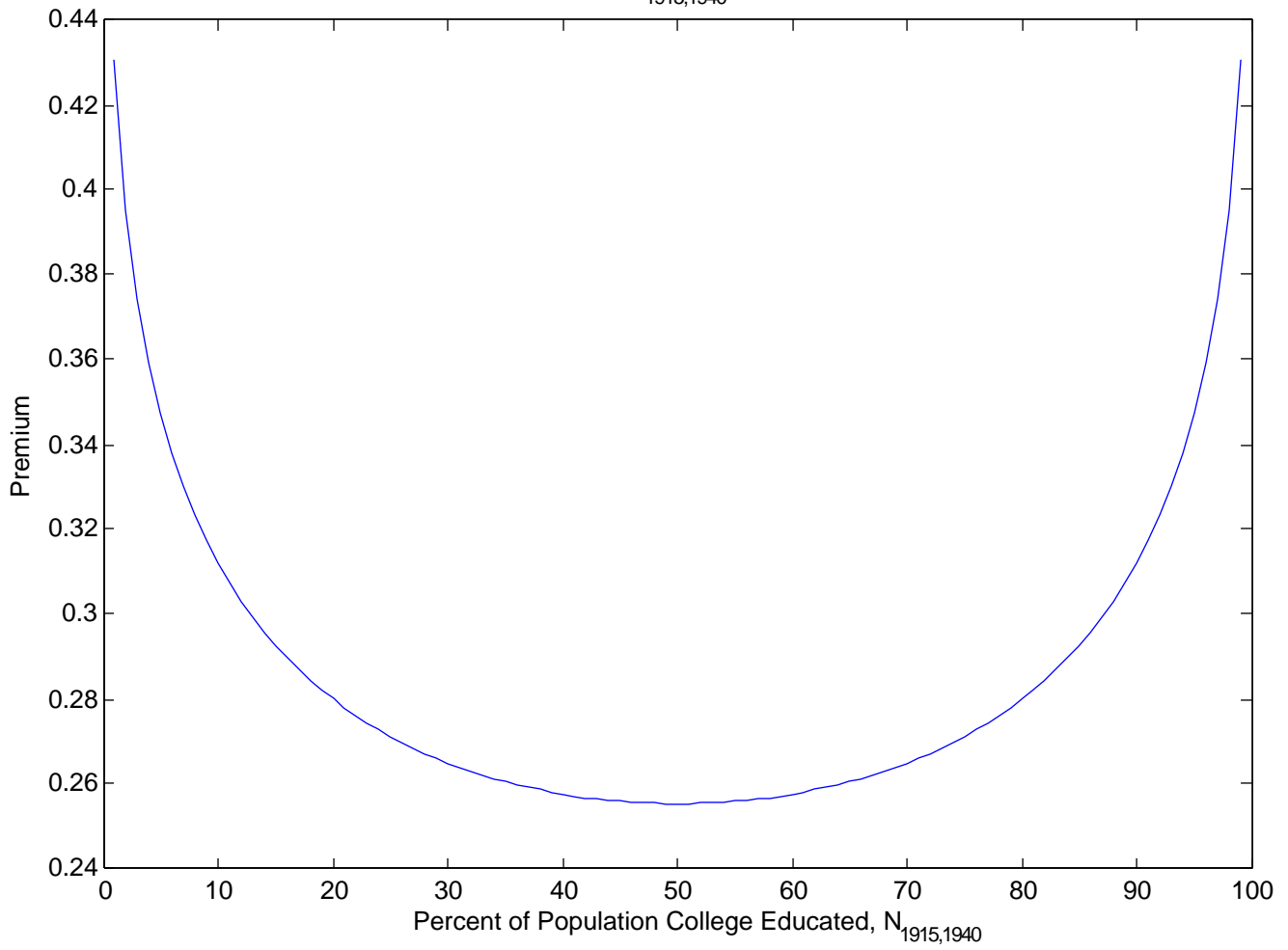


Figure 3: Counterfactual Education Premium as a Function of Percent of Population with a College Education due to Composition Effects, for 25 Year Olds in 1940, for $b_{1915,1940} = .9b_{1915,1940}^{OLS}$

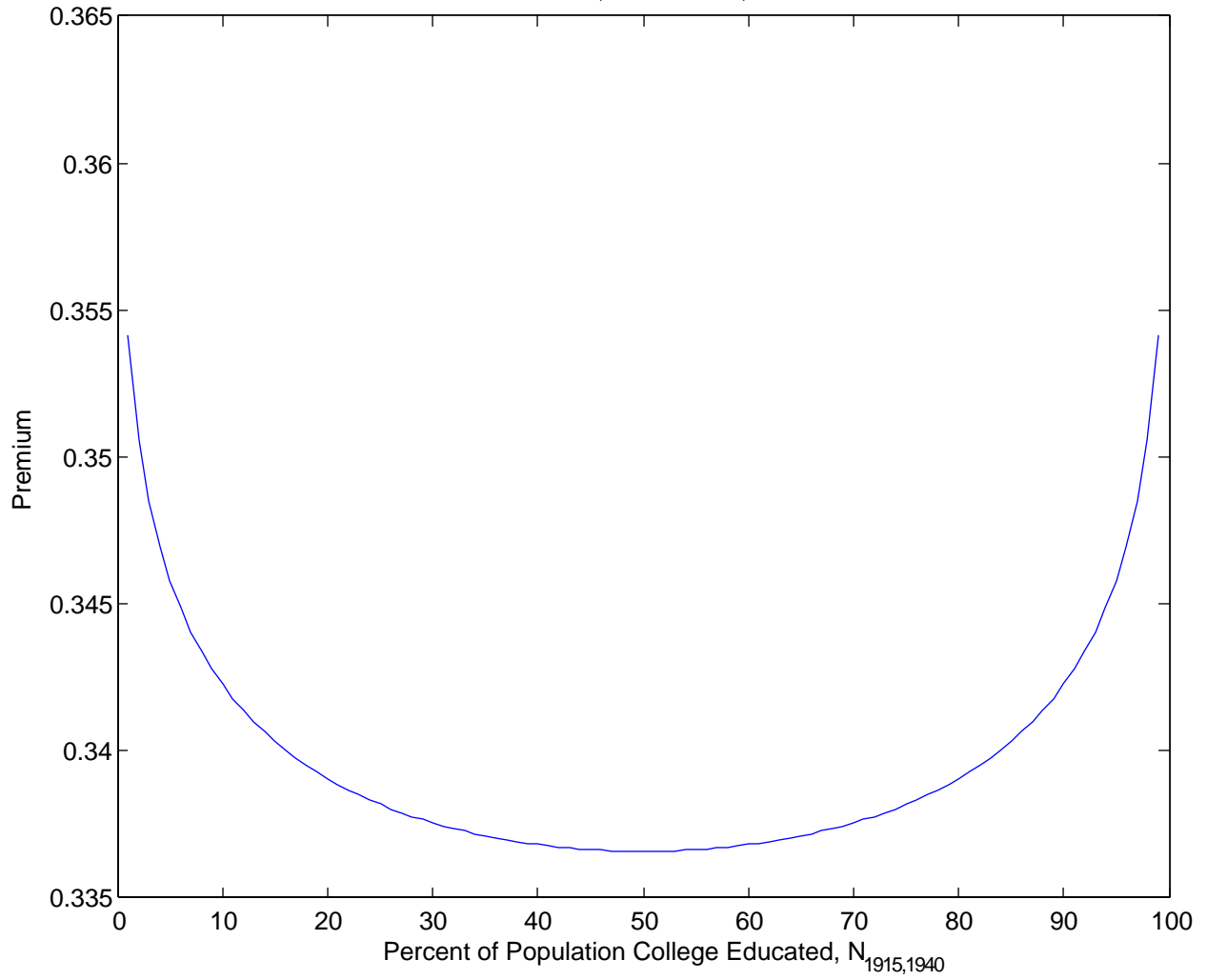


Figure 4: Percent College Educated by Cohort, N_{ct}

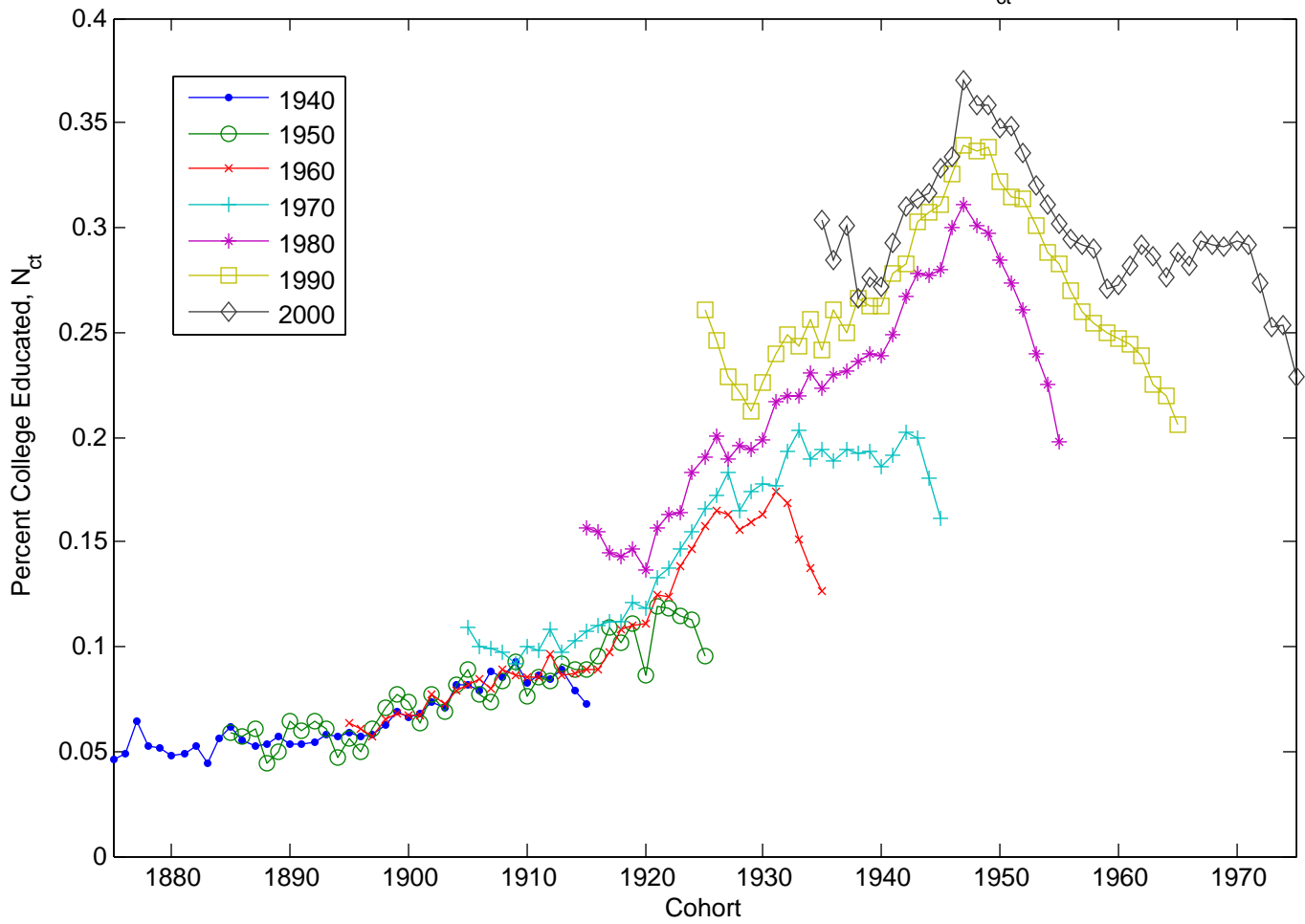


Figure 5: Standard Deviations of Education Related Ability by Age, 1940–1970, $b = 0$, for College

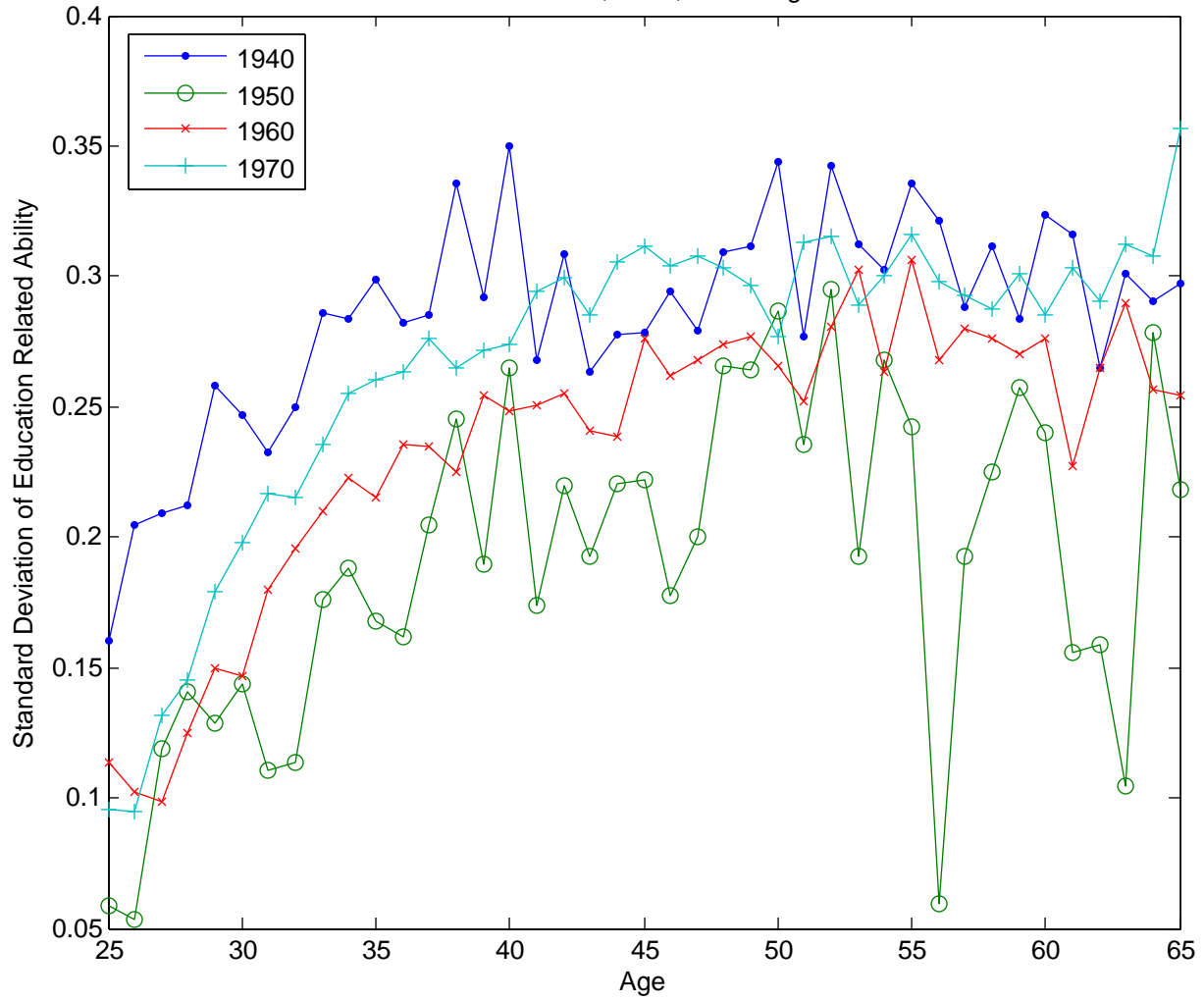


Figure 6: Standard Deviations of Education Related Ability by Age, 1980–2000, $b = 0$, for College

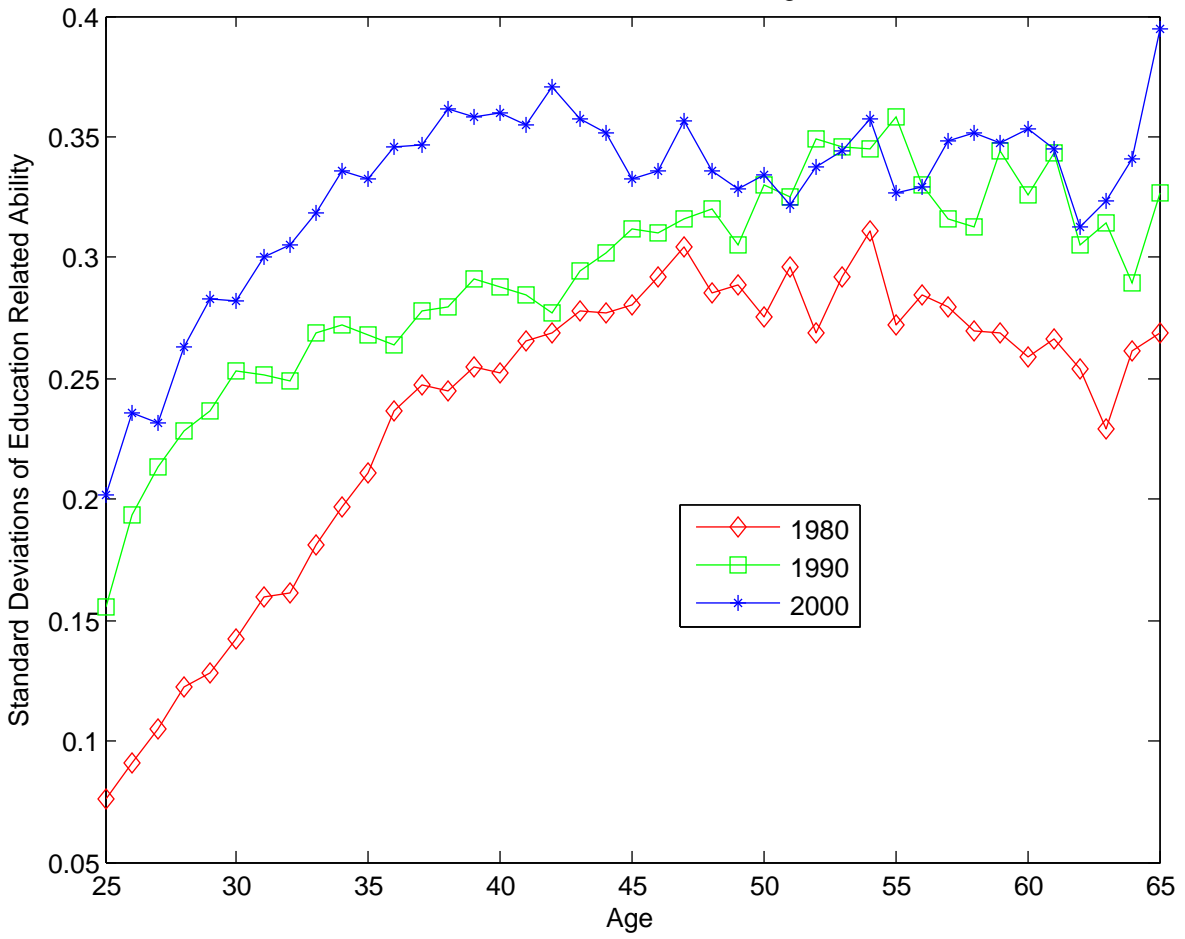


Figure 7: Mean Log Wages, w_{ct} 1940–1970, by Age = $c - t$

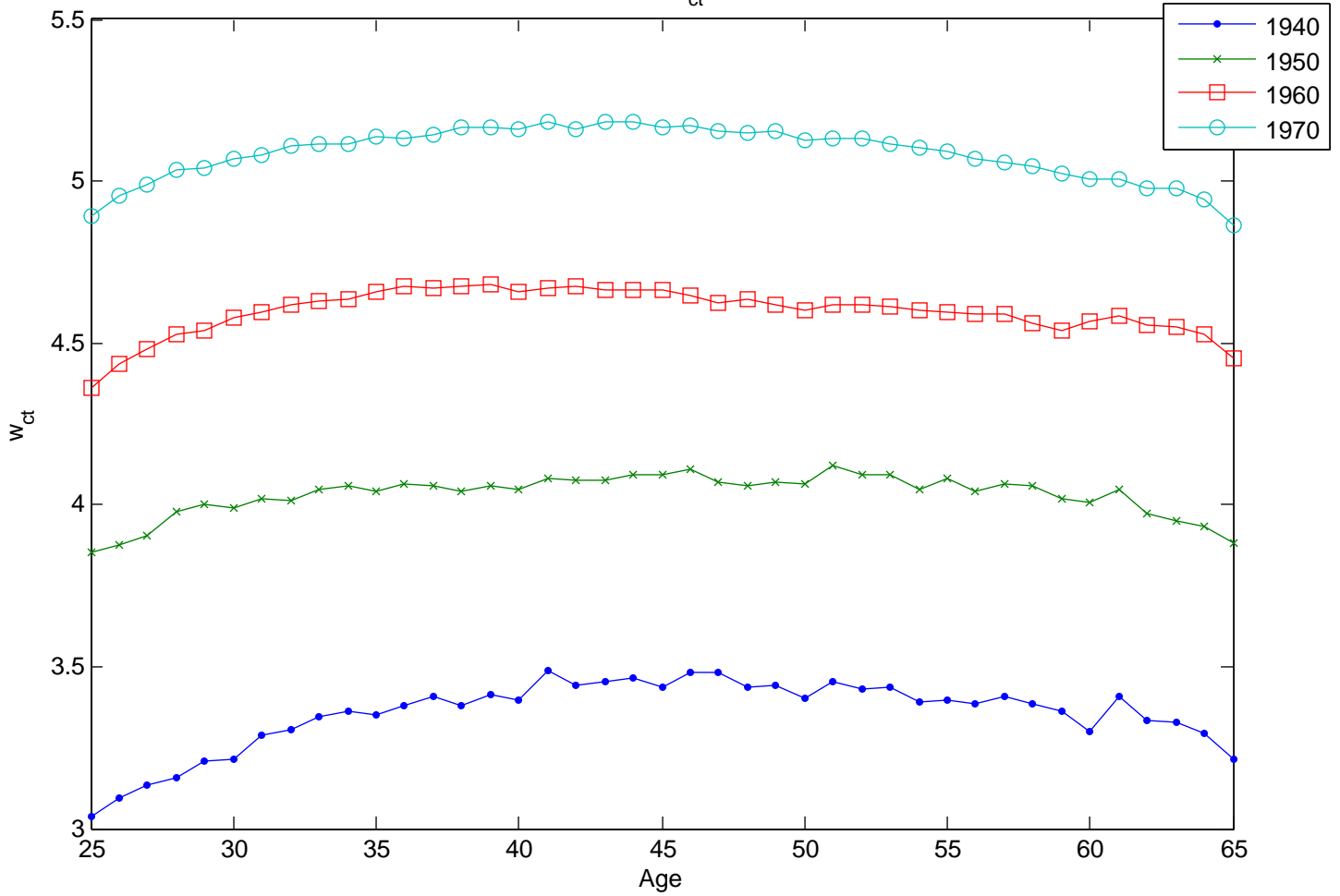


Figure 8: Mean Log Wages, w_{ct} 1980–2000, by Age = c - t

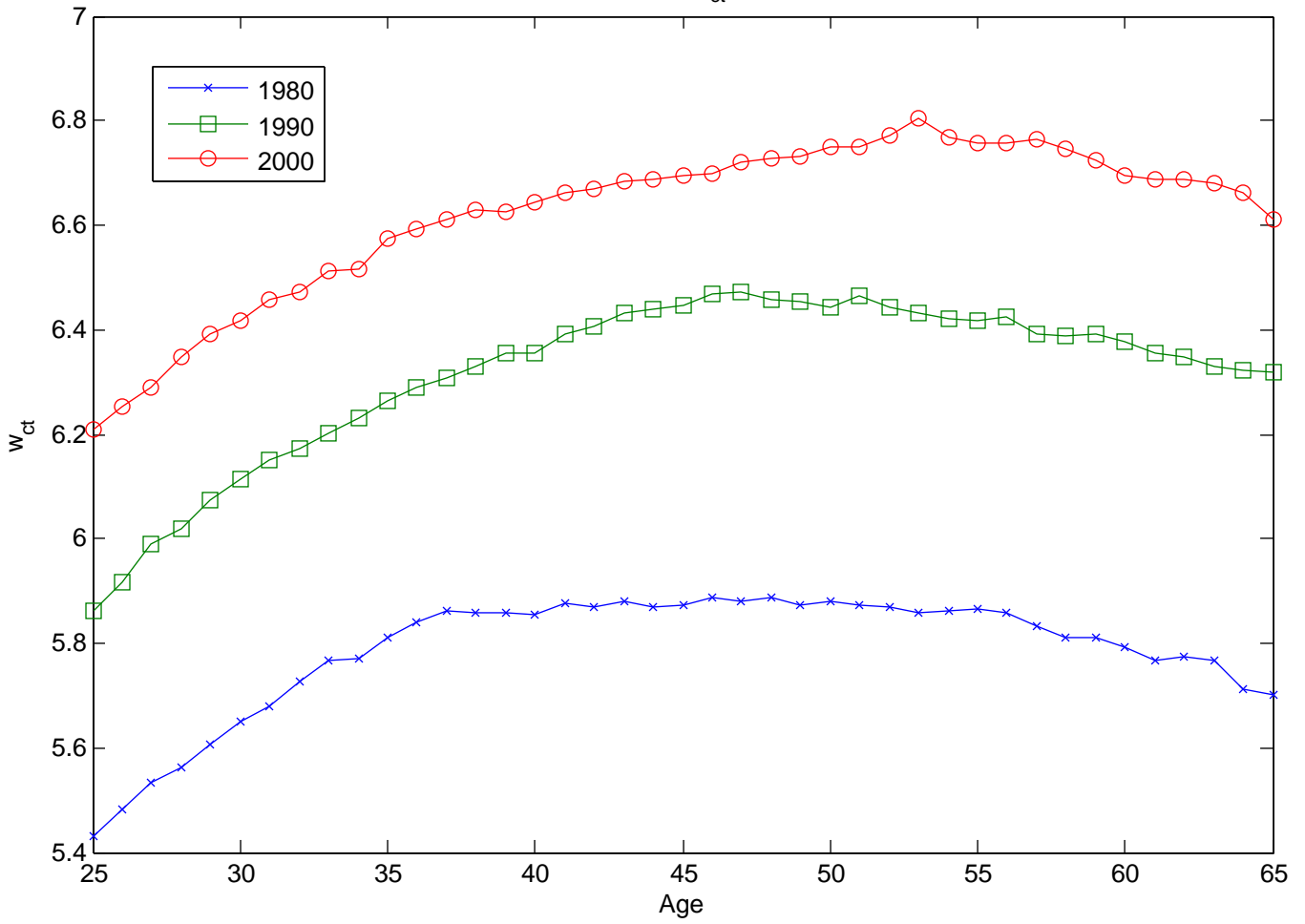


Figure 9: Direct Effect of College on Wages, for $b_{ct} = .5b_{ct}^{OLS}$,
1940–1970, by Age = c - t

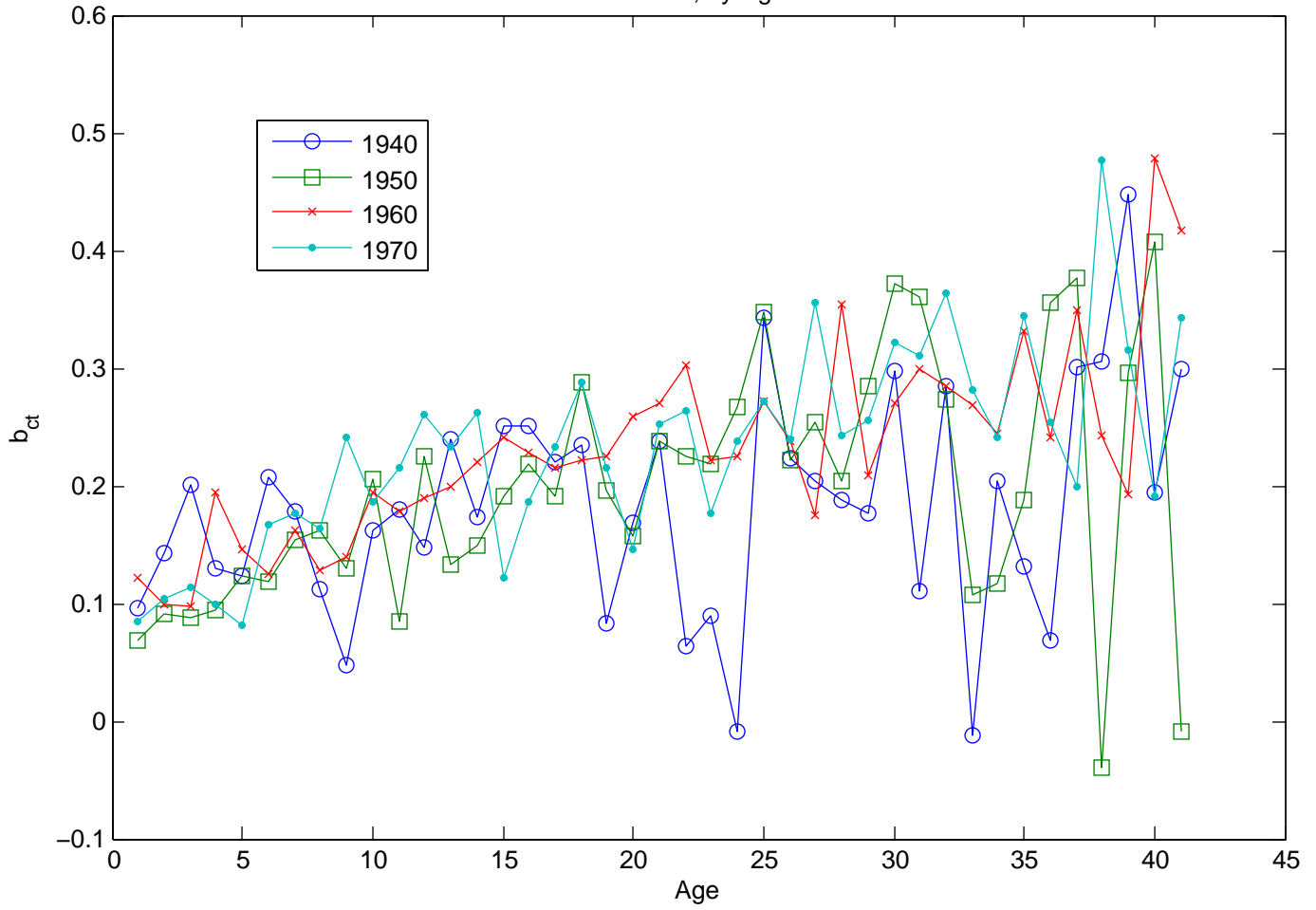


Figure 10: Direct Effect of College on Wages, for $b_{ct} = .5b_{ct}^{OLS}$,
1980–2000, by Age = c - t

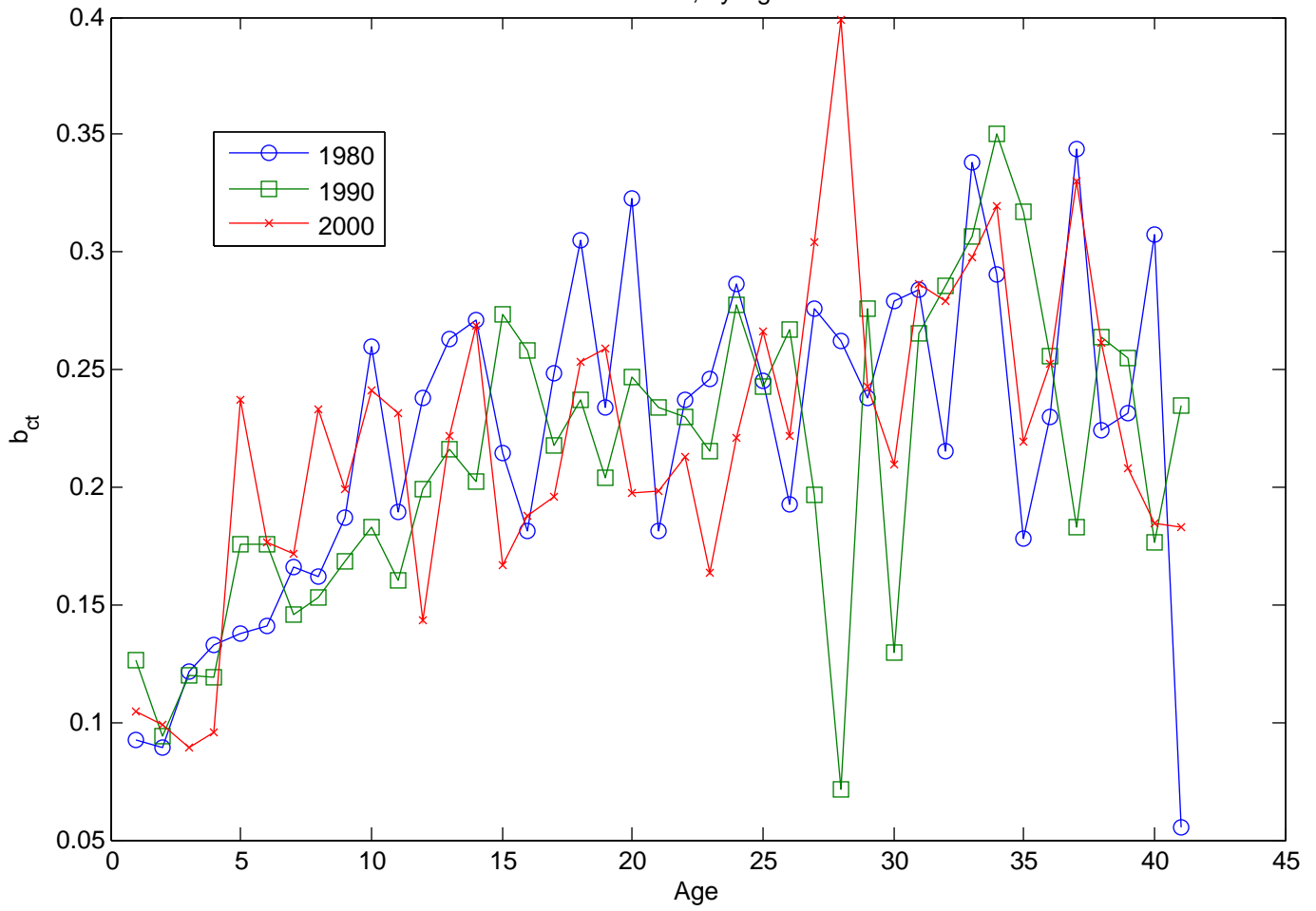


Figure 11: Mean College Correlated Ability, μ_{act} , for $b_{ct} = .5b_{ct}^{OLS}$, 1940–1970, by Age = c - t

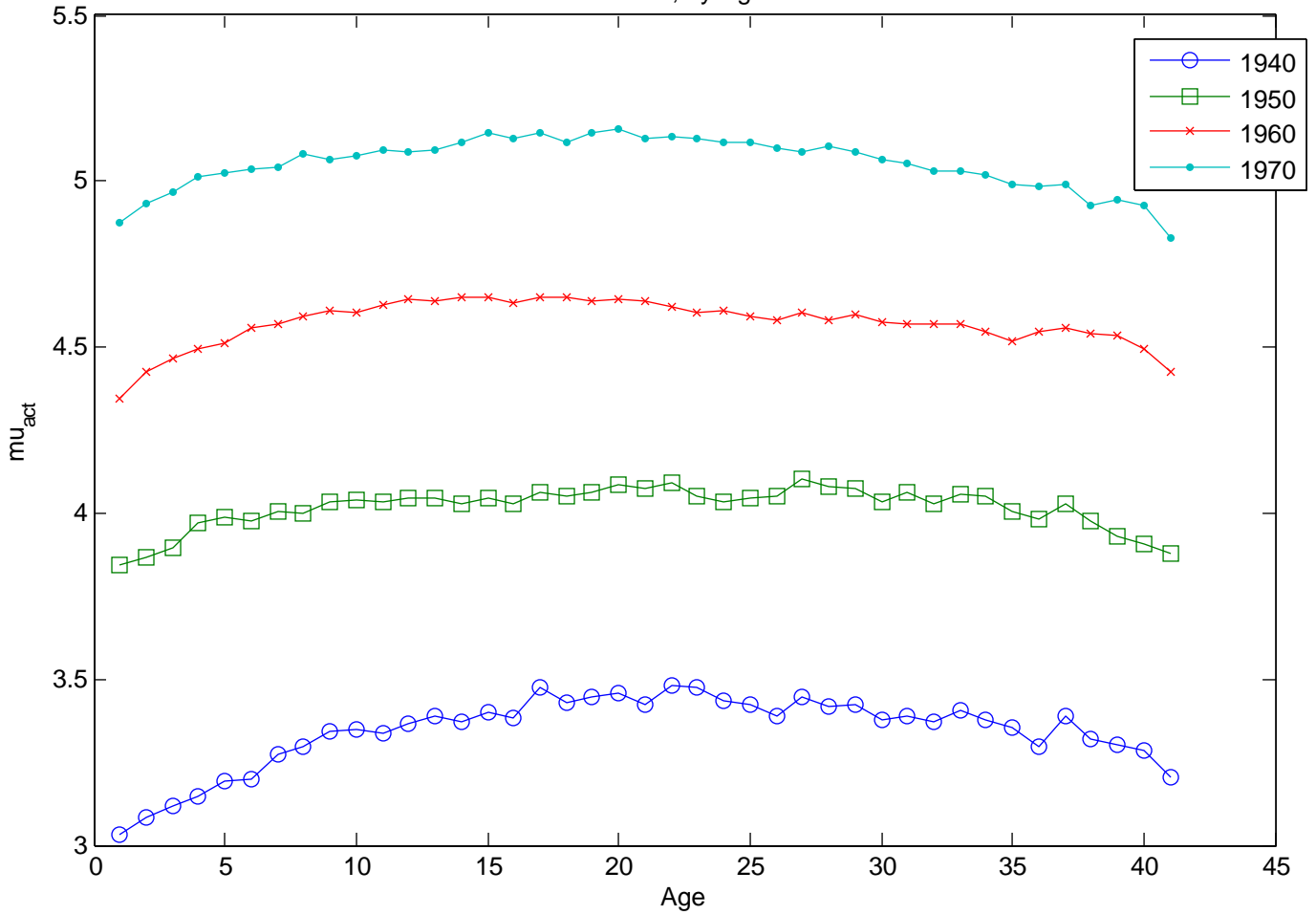


Figure 12: Mean College Correlated Ability, μ_{act} , for $b_{ct} = .5b_{ct}^{OLS}$, 1980–2000, by Age = $c - t$

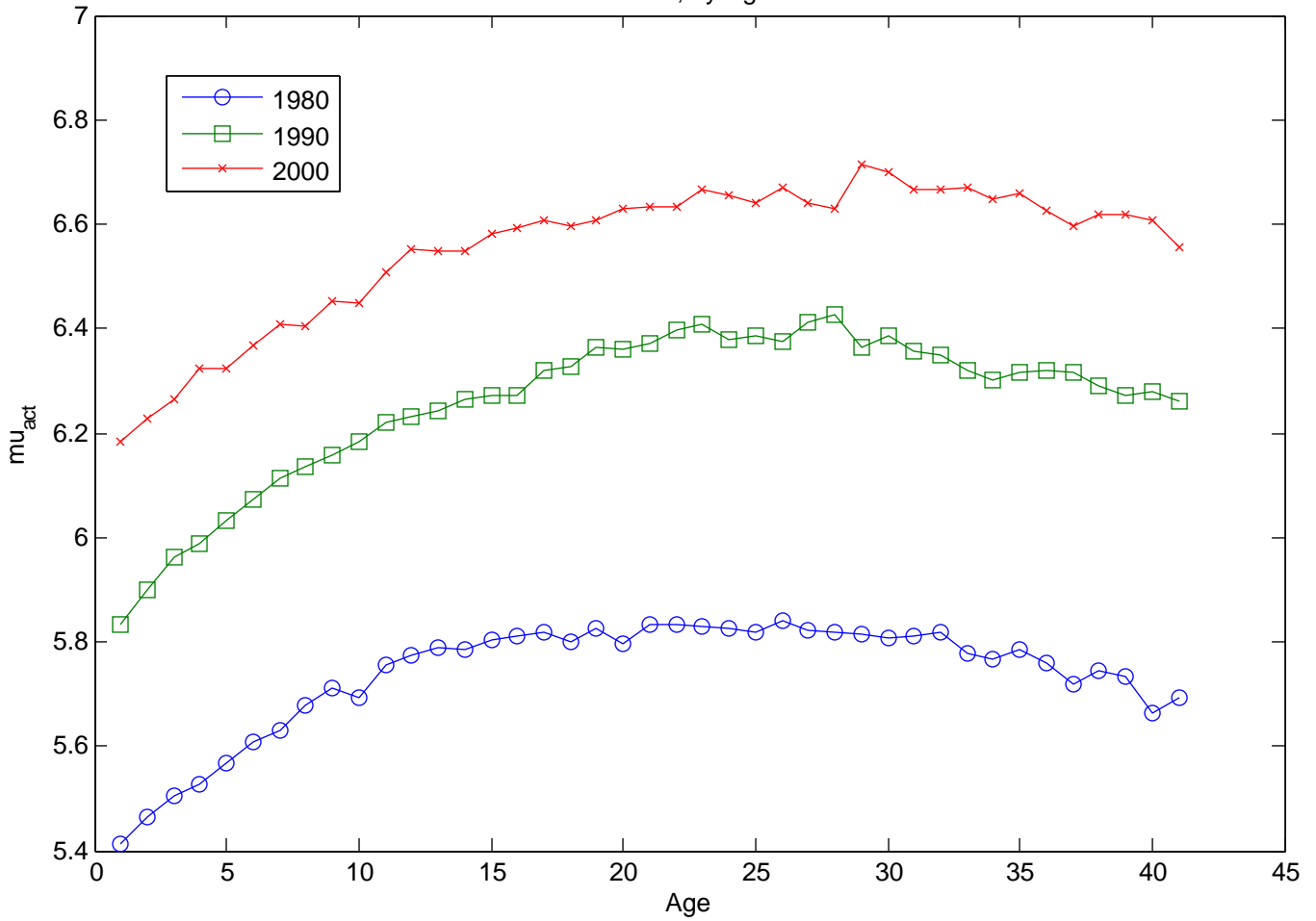


Figure 13: Direct Effect of College on Wages, for $b_{ct} = .9b_{ct}^{OLS}$,
1940–1970, by Age = c - t

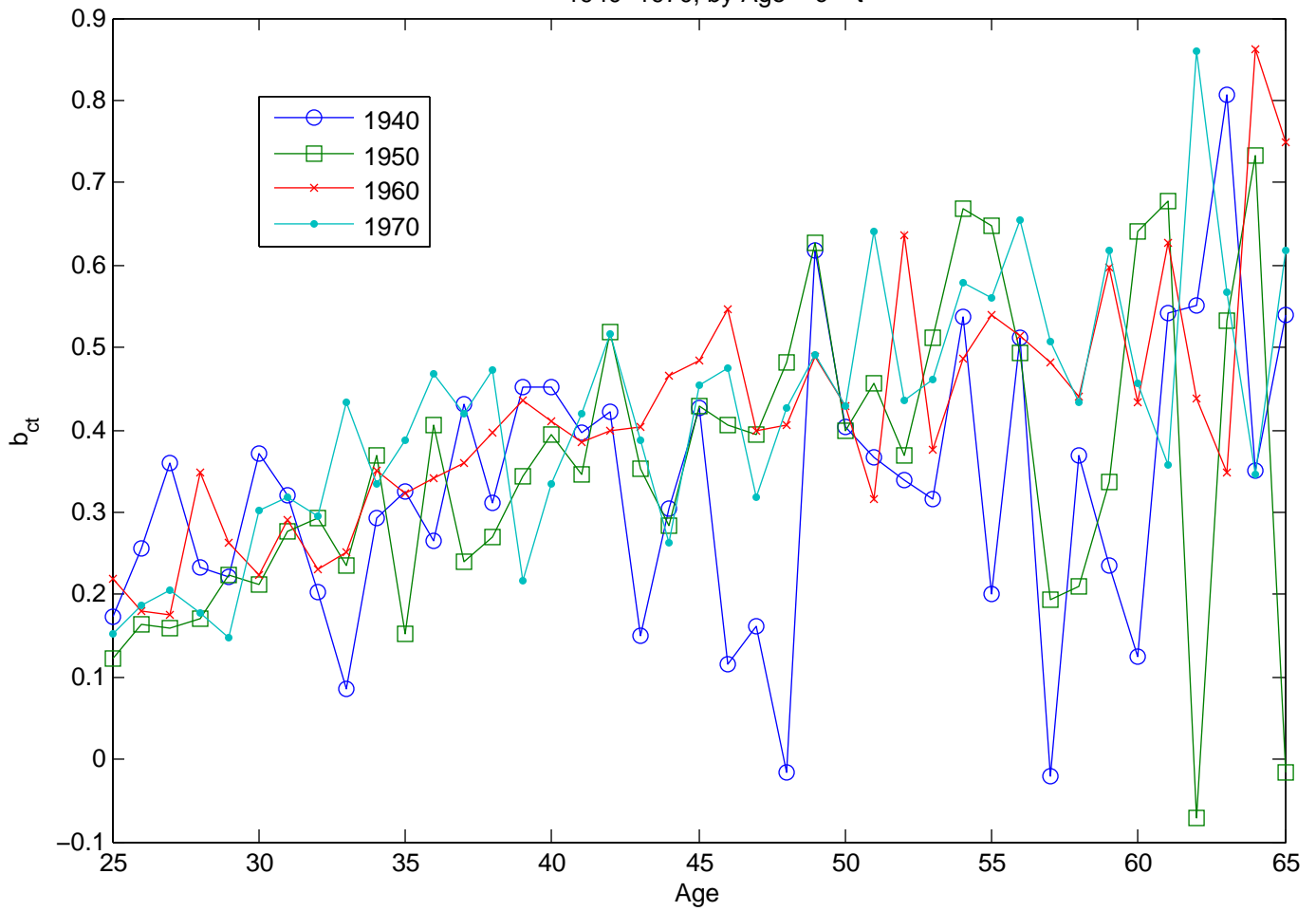


Figure 14: Direct Effect of College on Wages, for $b_{ct} = .9b_{ct}^{OLS}$,
1980–2000, by Age = c - t

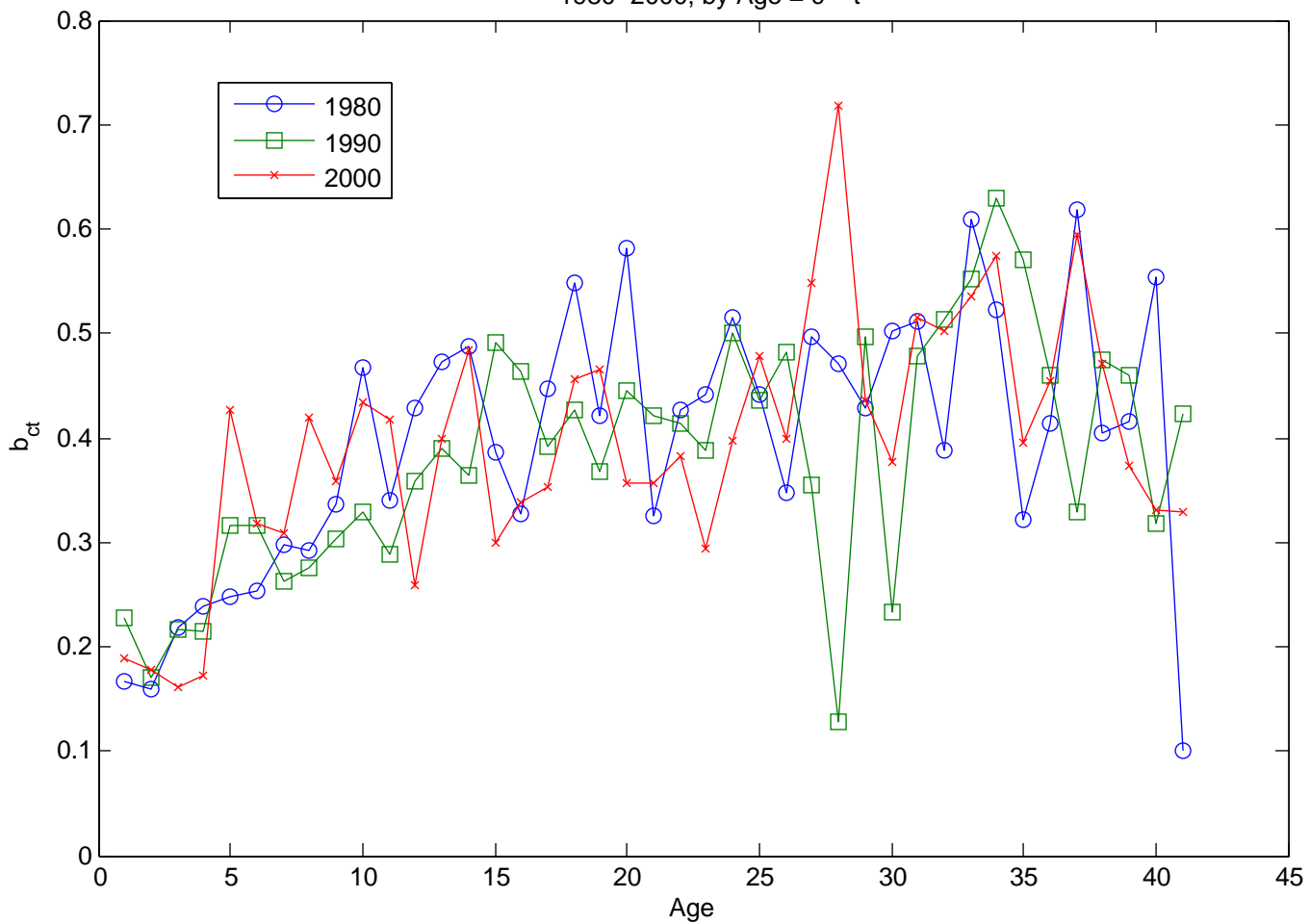


Figure 15: Mean College Correlated Ability, μ_{act} for $b_{ct} = .9b_{ct}^{OLS}$, 1940–1970, by Age = $c - t$

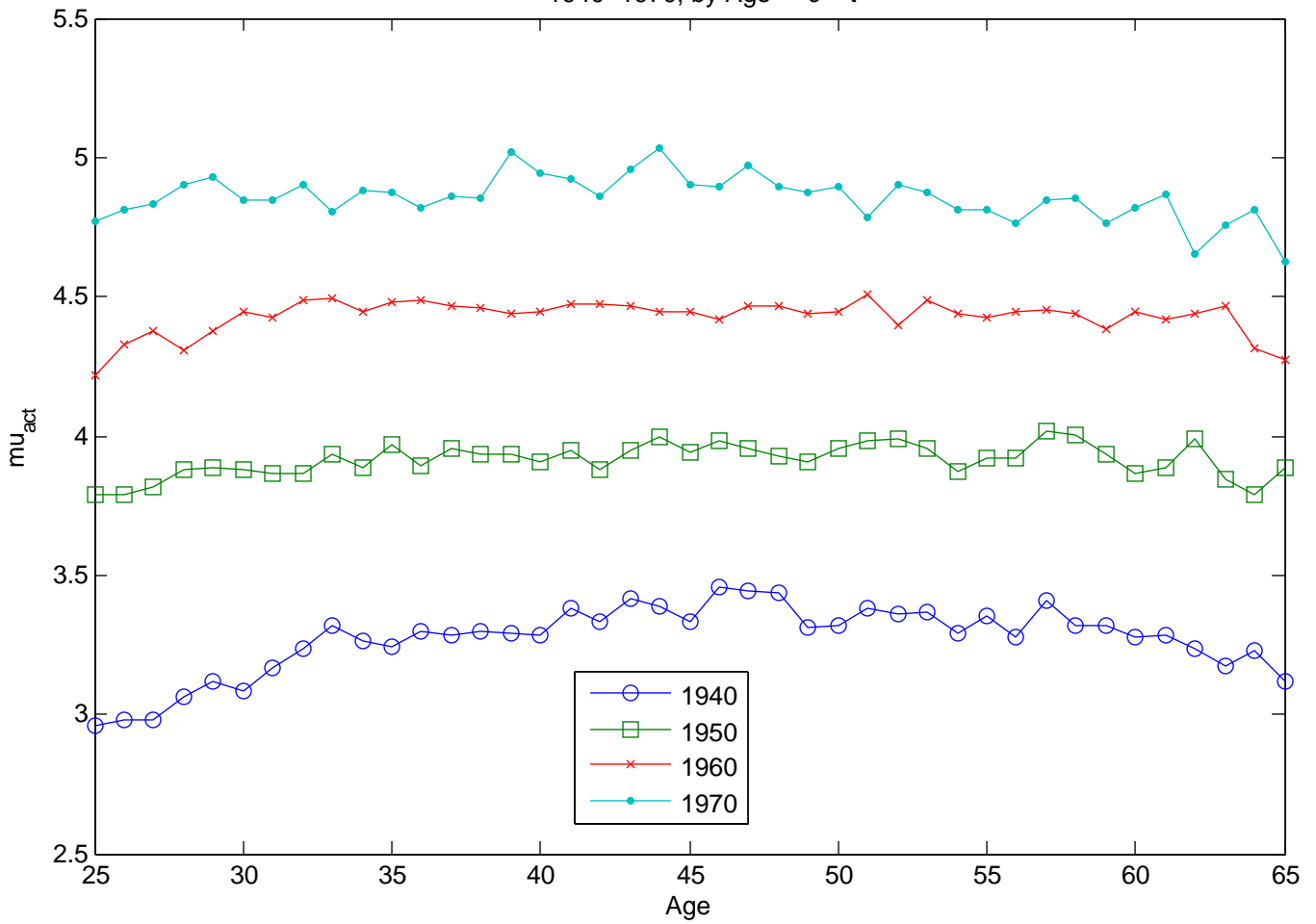


Figure 16: Mean College Correlated Ability, μ_{act} for $b_{ct} = .9b_{ct}^{OLS}$, 1980–2000, by Age = $c - t$

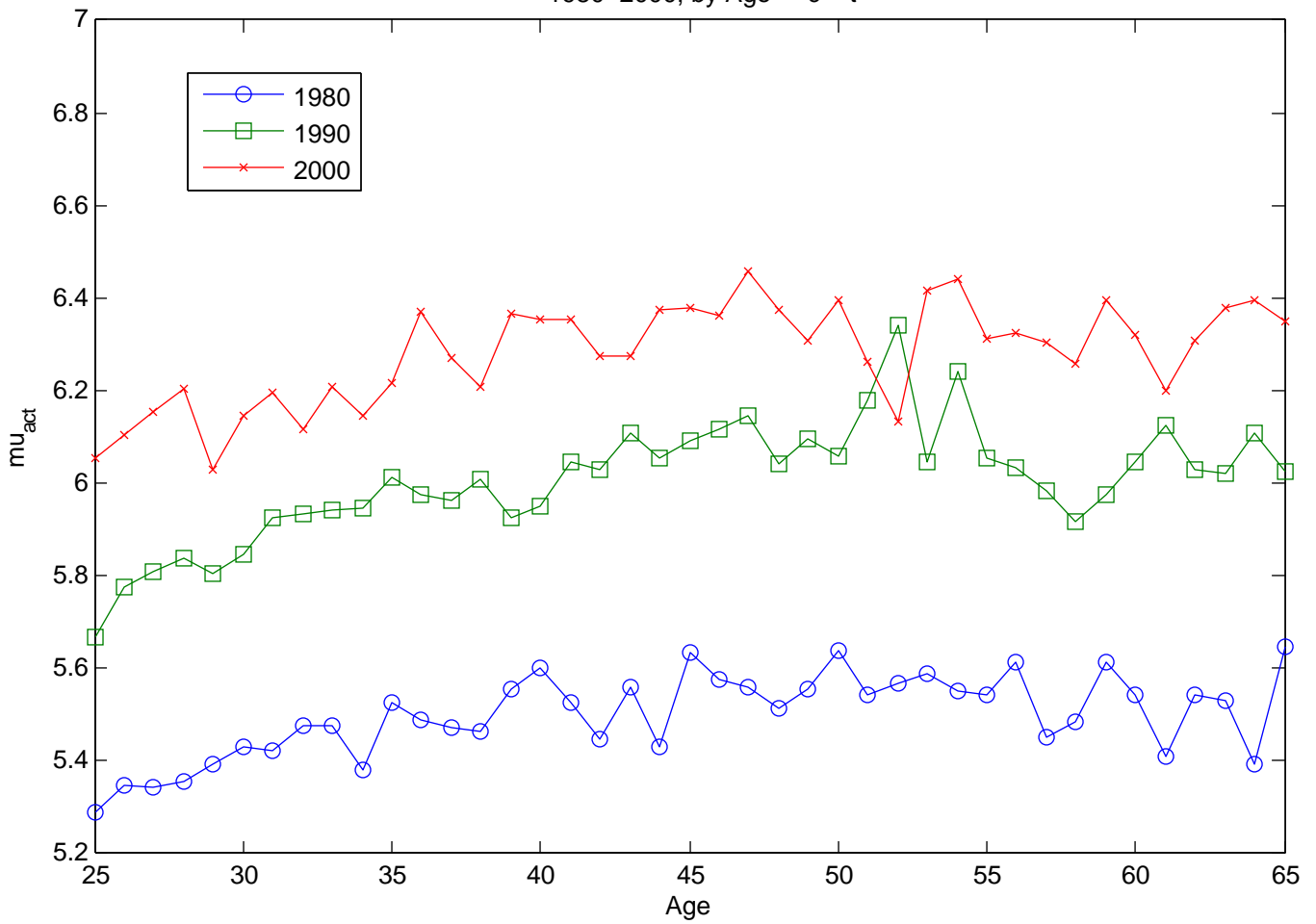


Figure 17: Percent High School Educated by Cohort, N_{ct}^{HS}

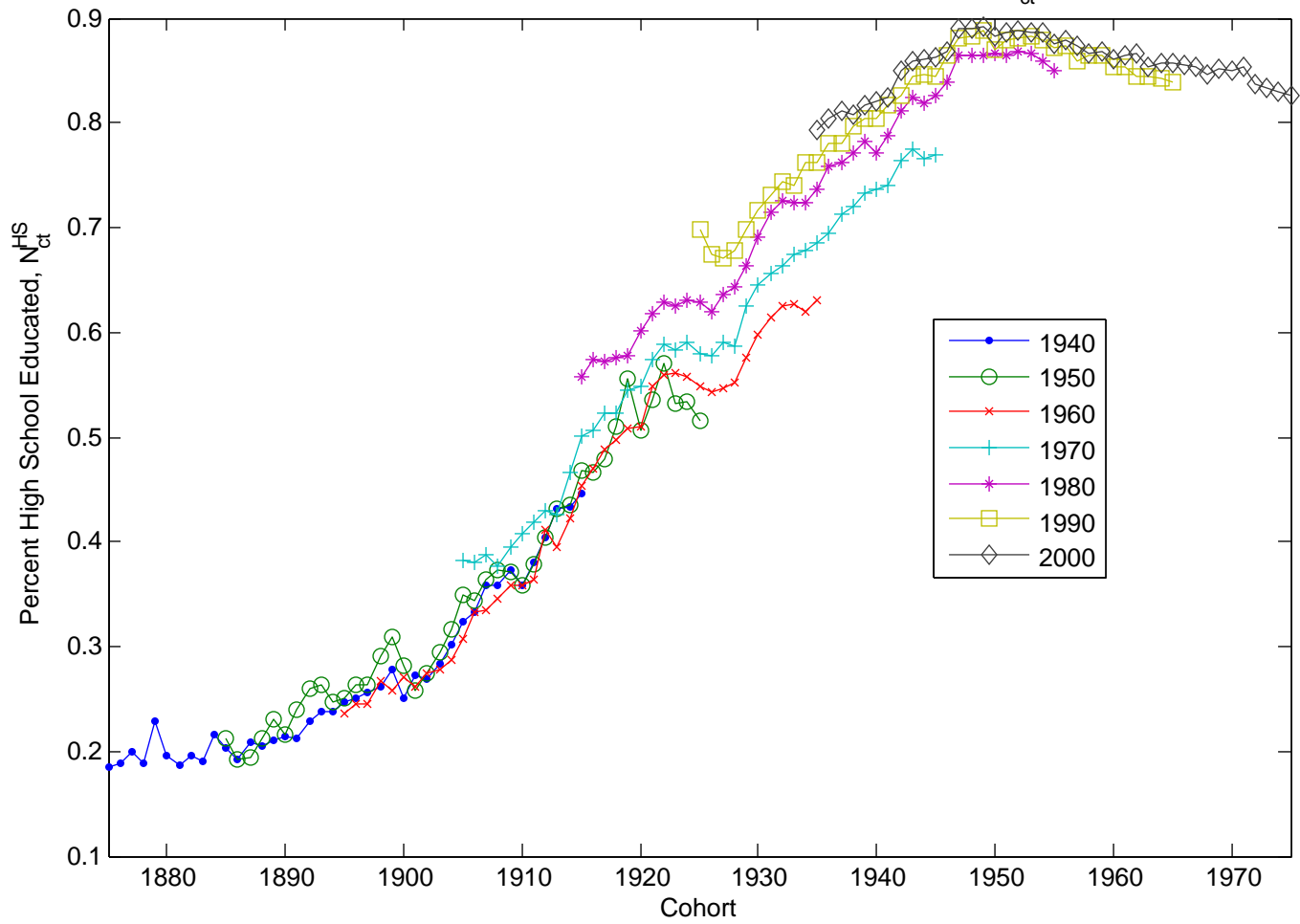


Figure 18: Standard Deviation of Education Related Ability by Age, 1940–1970, $b = 0$, for High School

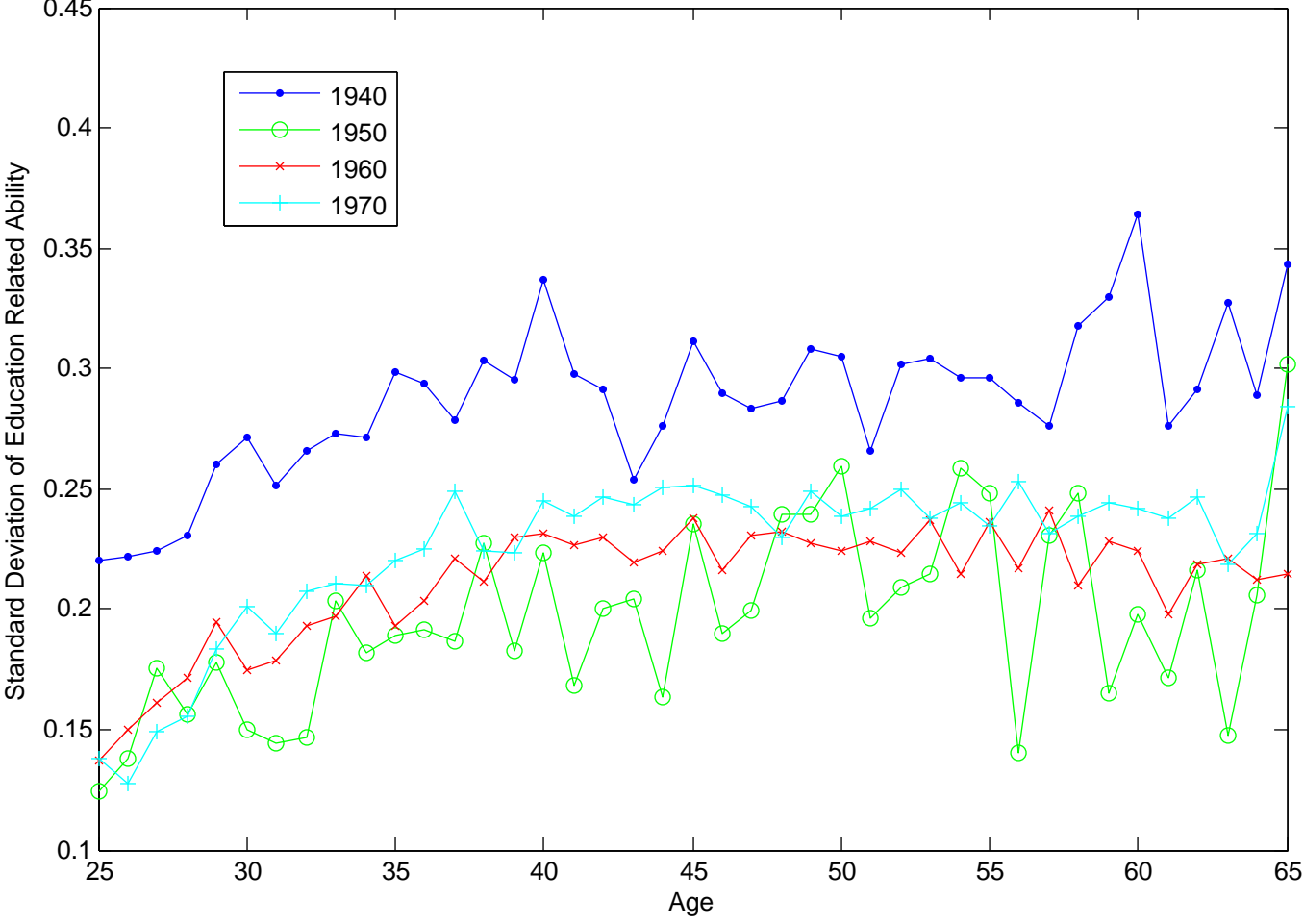


Figure 19: Standard Deviation of Education Related Ability by Age, 1980–2000, $b = 0$, for High School

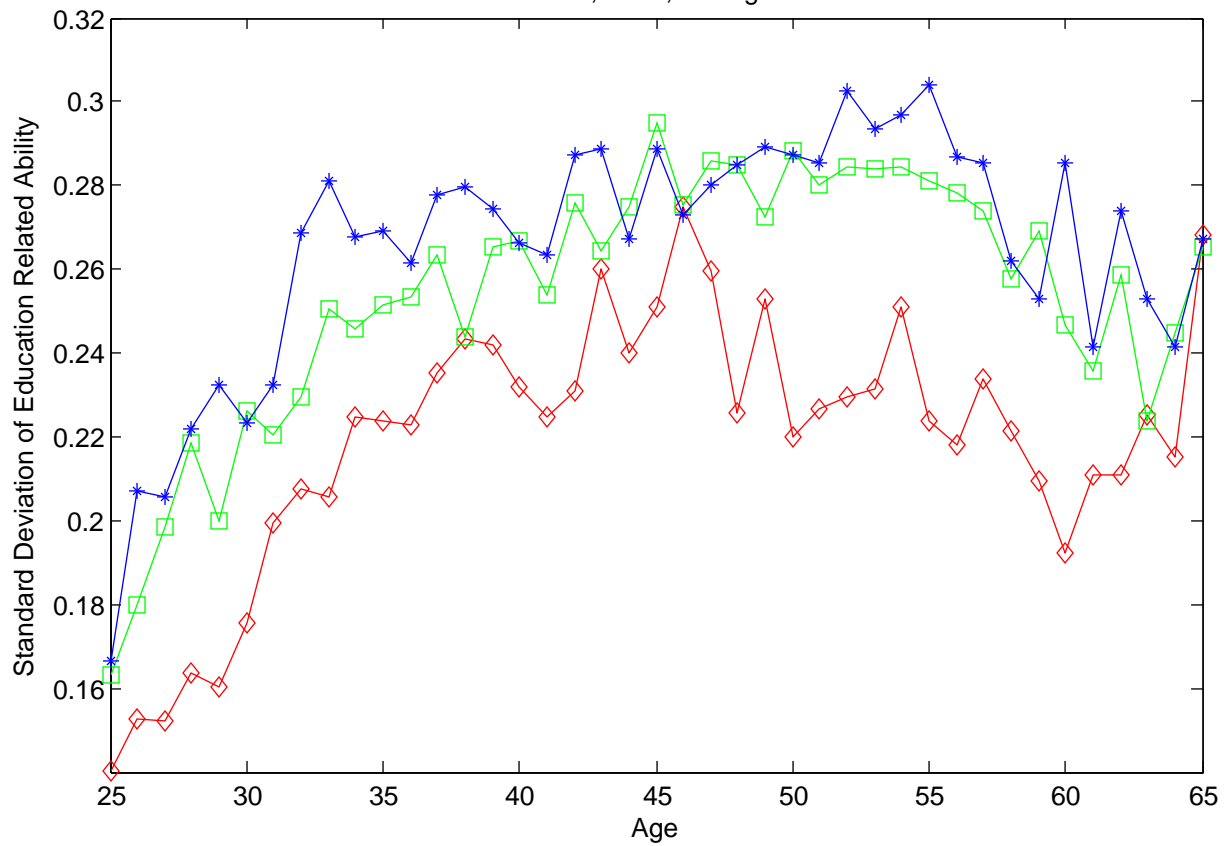


Figure 20: Standard Deviation of Education Related Ability by Age, 1940–1970, $b_{ct} = .5b_{ct}^{OLS}$, for High School

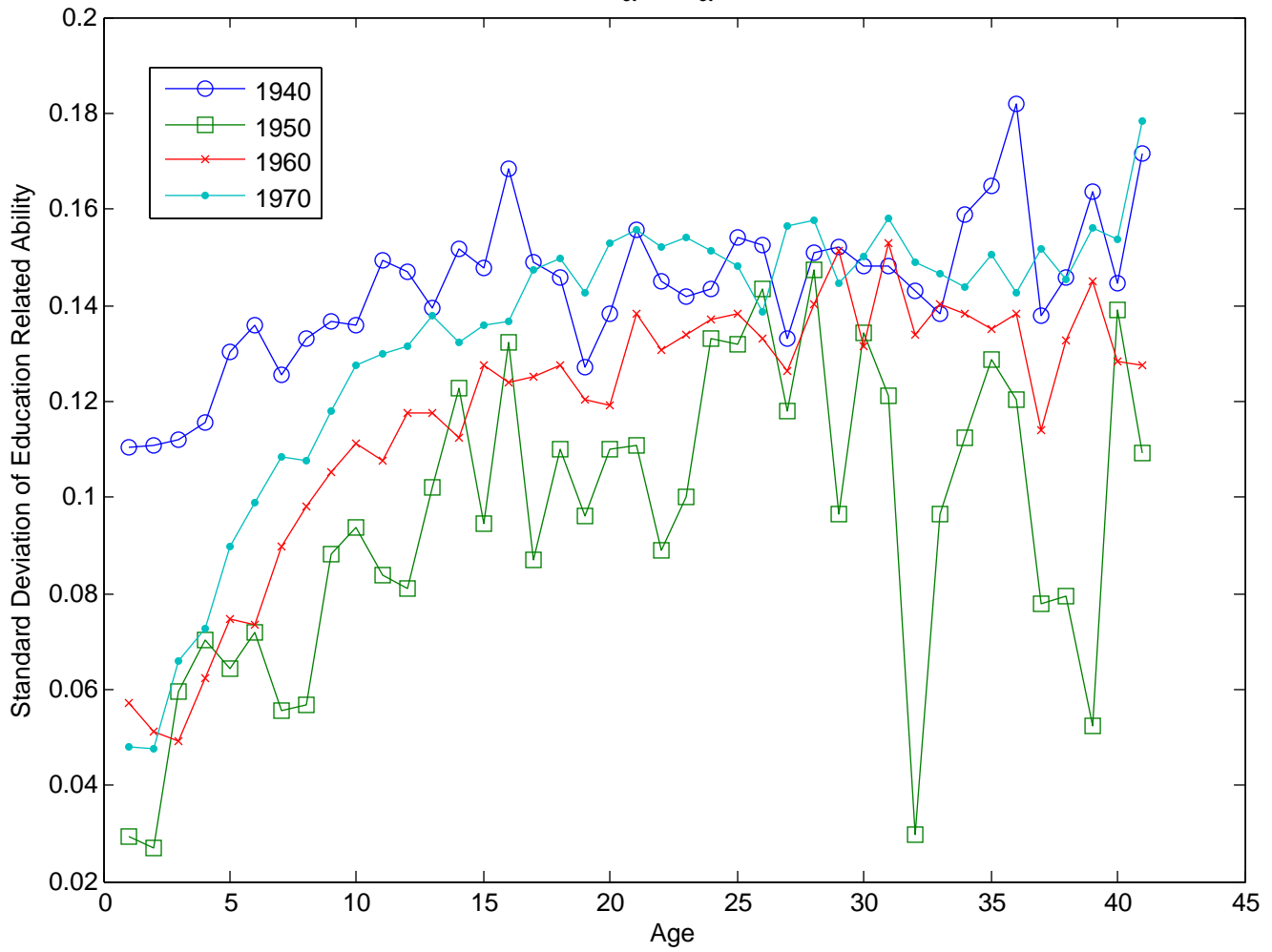


Figure 21: Standard Deviation of Education Related Ability by Age, 1980–2000, $b_{ct} = .5b_{ct}^{OLS}$, for High School

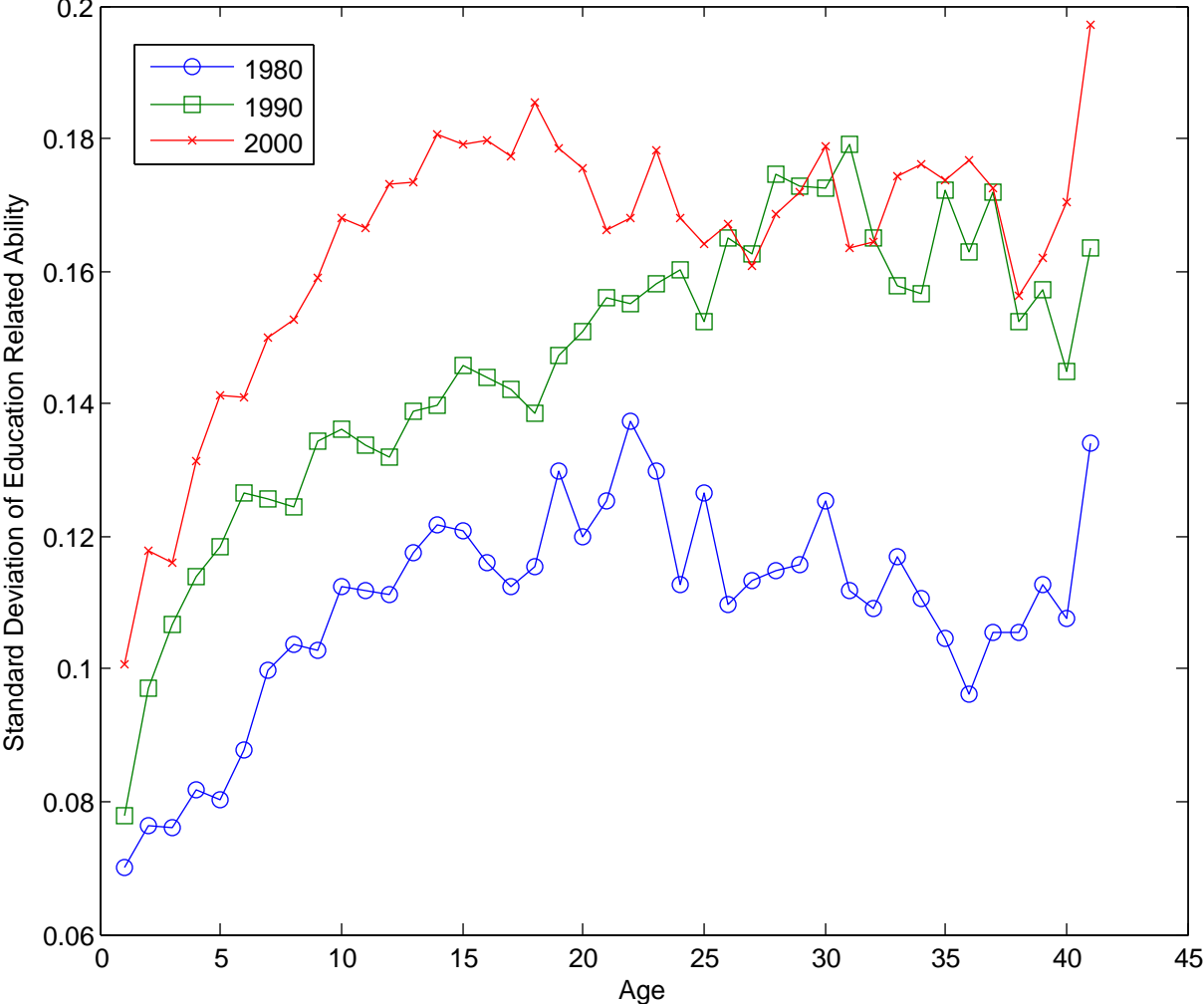


Figure 22: Direct Effect of High School on Wages, for $b_{ct} = .5b_{ct}^{OLS}$, 1940–1970, by Age = c - t

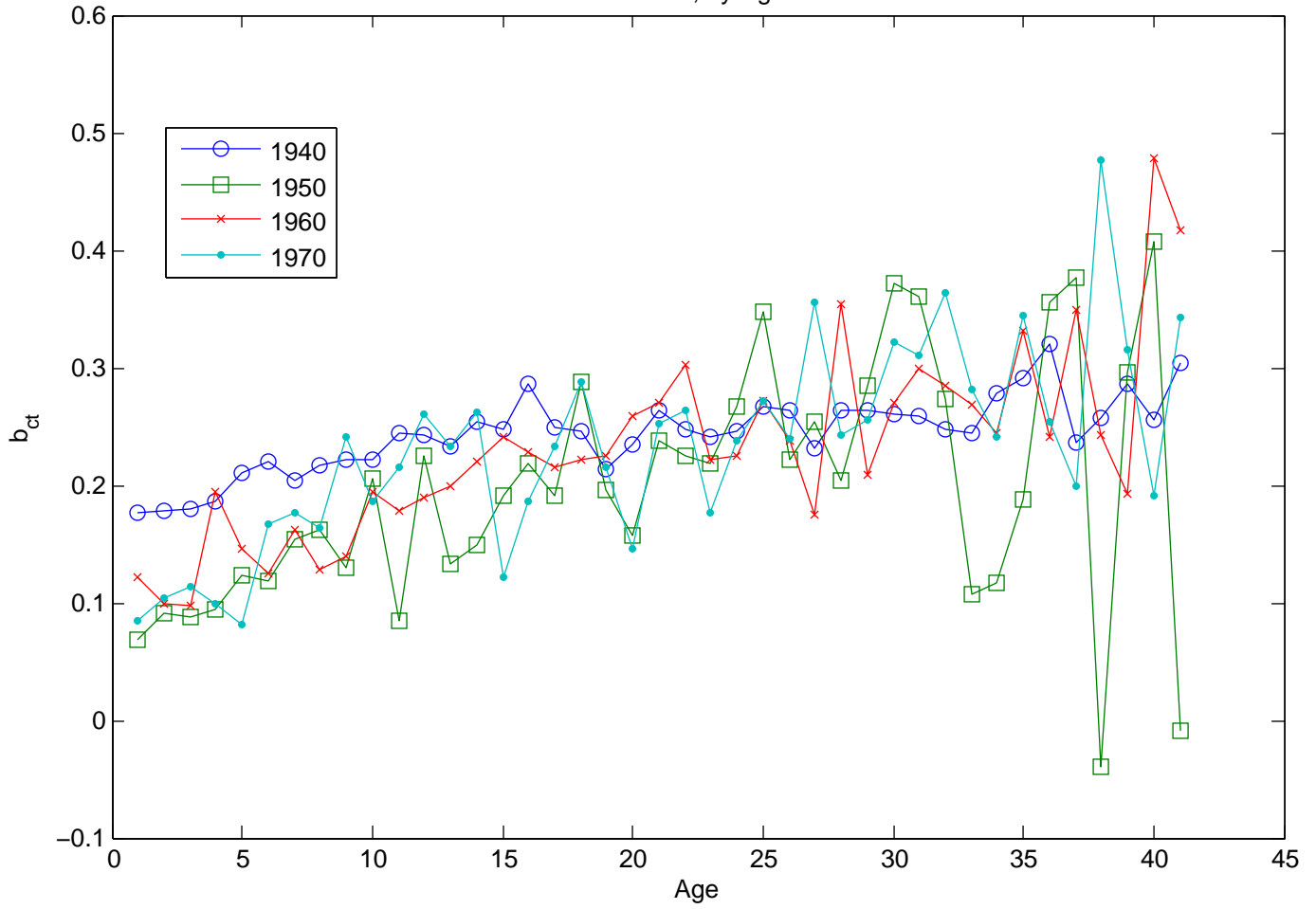


Figure 23: Direct Effect of High School on Wages, for $b_{ct} = .5b_{ct}^{OLS}$, 1980–2000, by Age = c - t

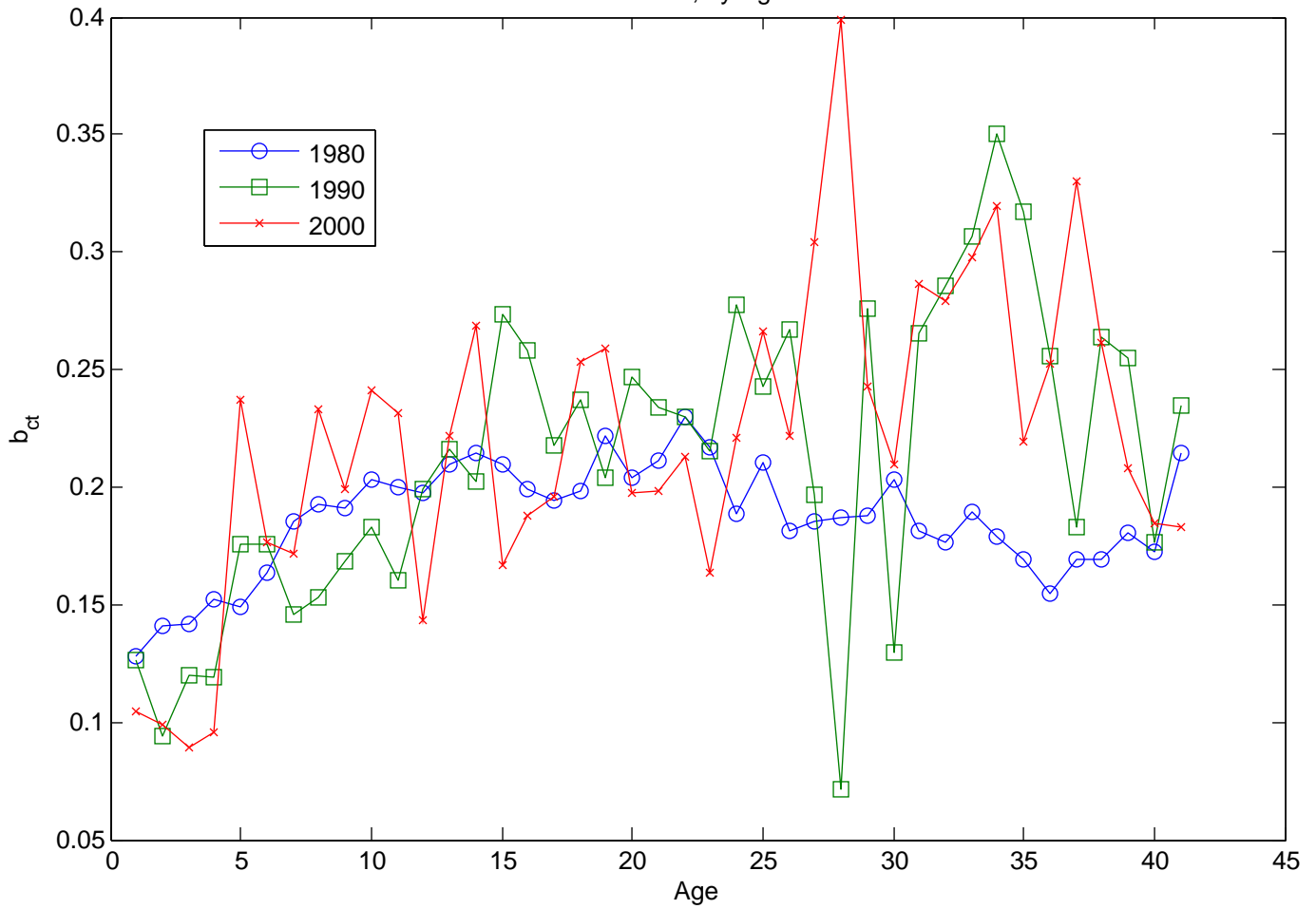


Figure 24: Mean High School Correlated Ability, μ_{act} , for $b_{ct} = .5b_{ct}^{OLS}$, 1940–1970, by Age = $c - t$

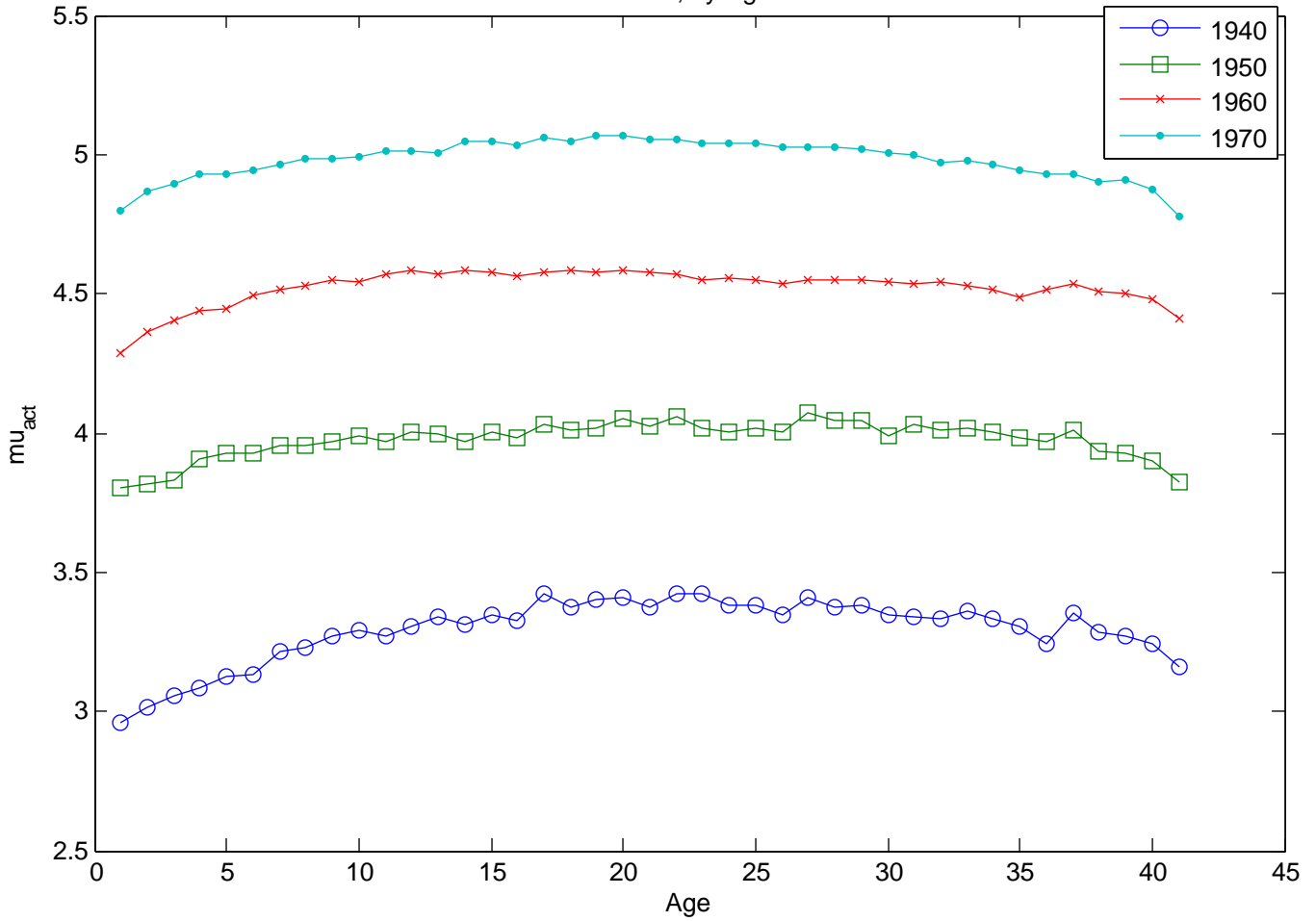


Figure 25: Mean High School Correlated Ability, μ_{act} , for $b_{ct} = .5b_{ct}^{OLS}$, 1980–2000, by Age = $c - t$

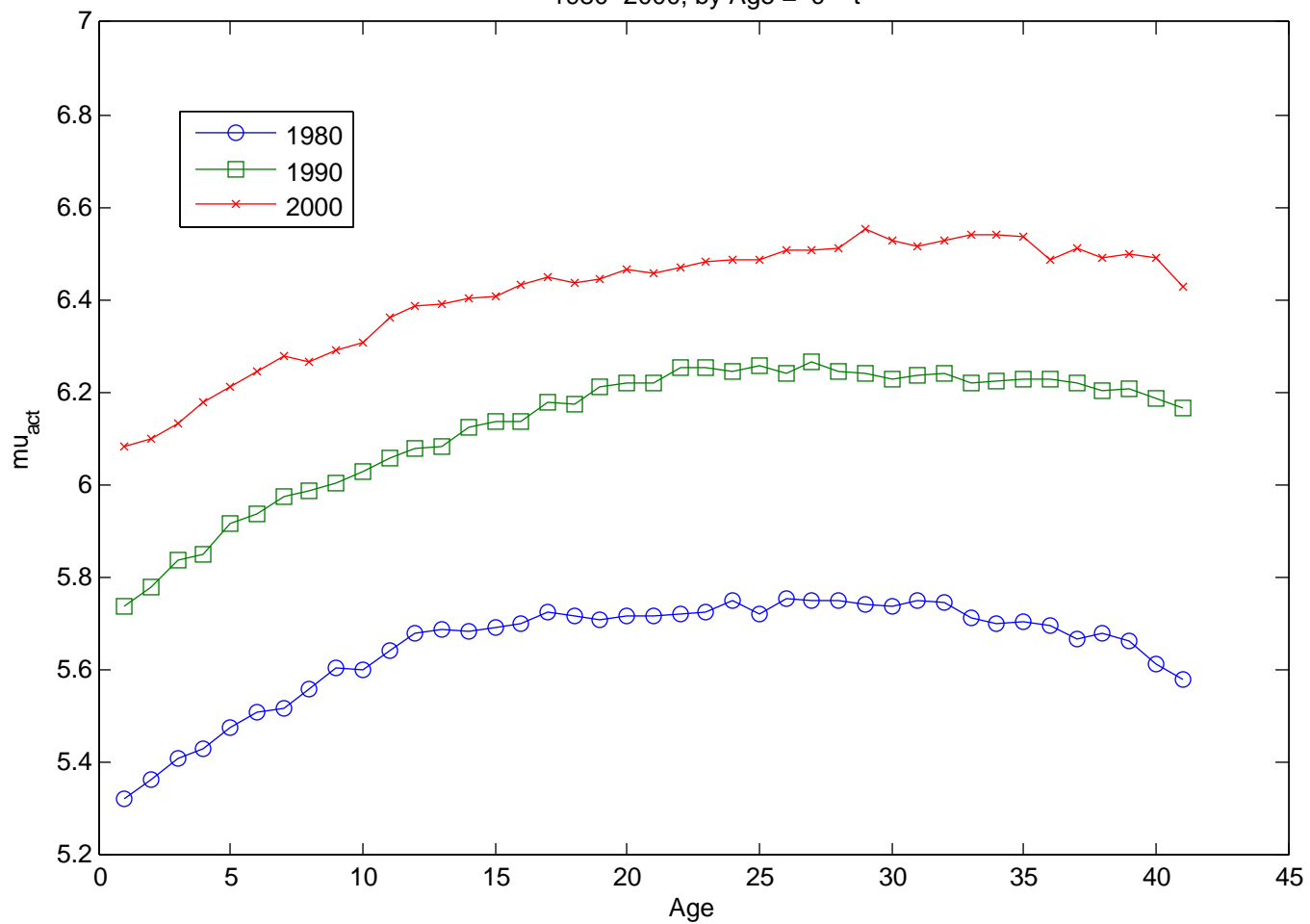


Figure 26: Standard Deviations of Education Correlated Ability by Age, 1940-1970, $b_{ct} = .9b_{ct}^{OLS}$, for High School

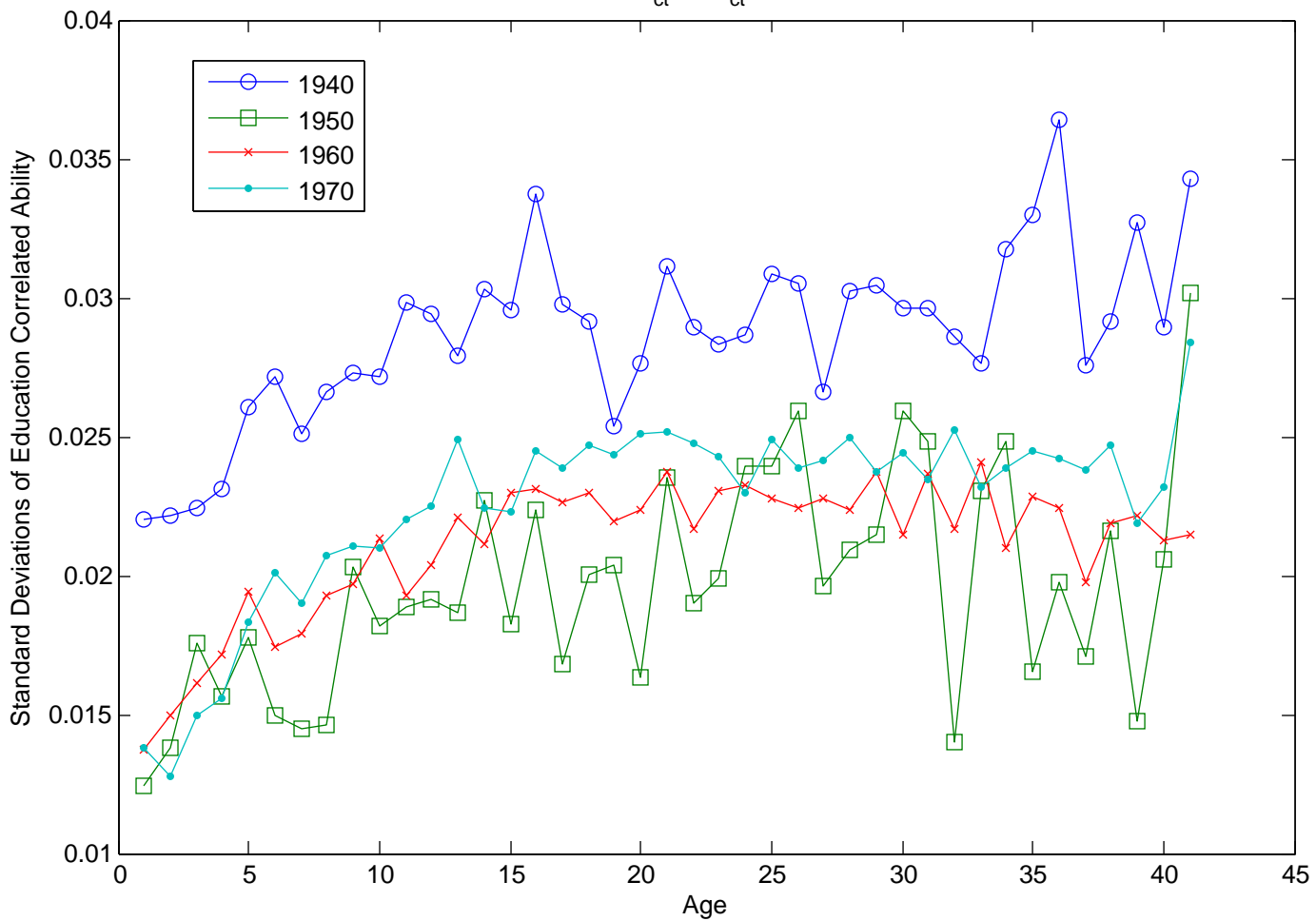


Figure 27: Standard Deviations of Education Correlated Ability by Age, 1980–2000, $b_{ct} = .9b_{ct}^{OLS}$, for High School

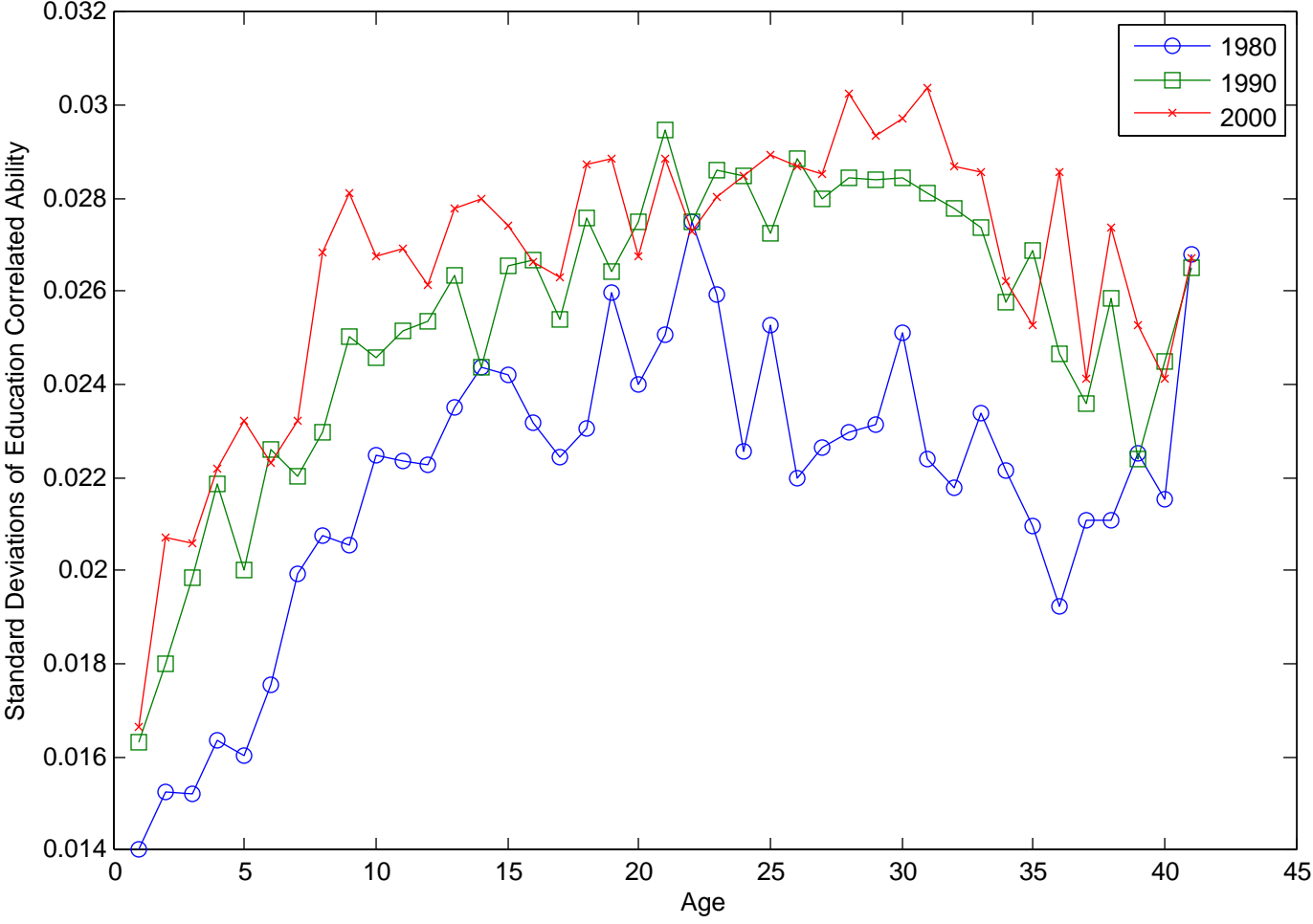


Figure 28: Direct Effect of High School on Wages, $b_{ct} = .9b_{ct}^{OLS}$,
1940–1970, by Age = c - t

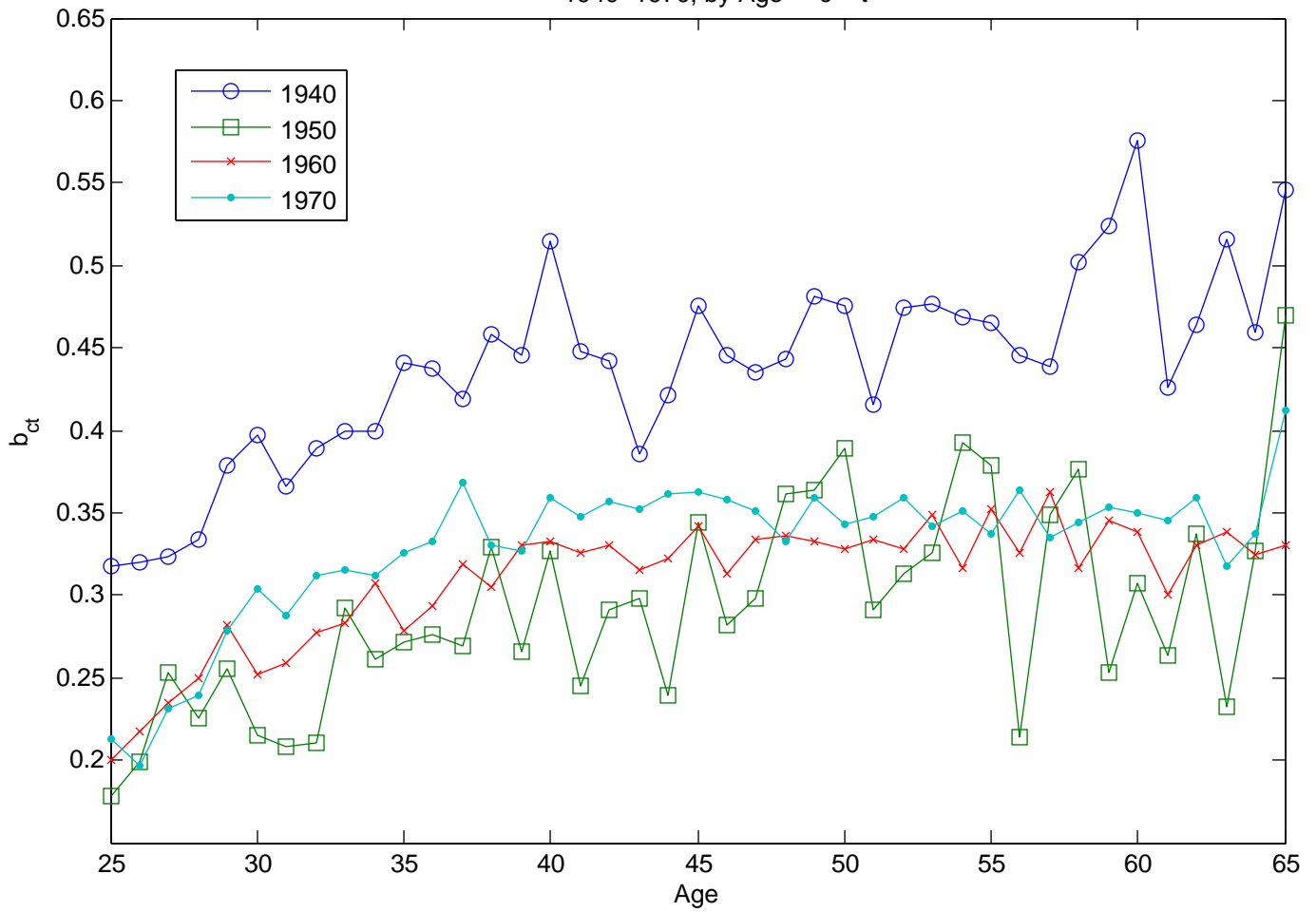


Figure 29: Direct Effect of High School on Wages, $b_{ct} = .9b_{ct}^{OLS}$,
1980–2000, by Age = c - t

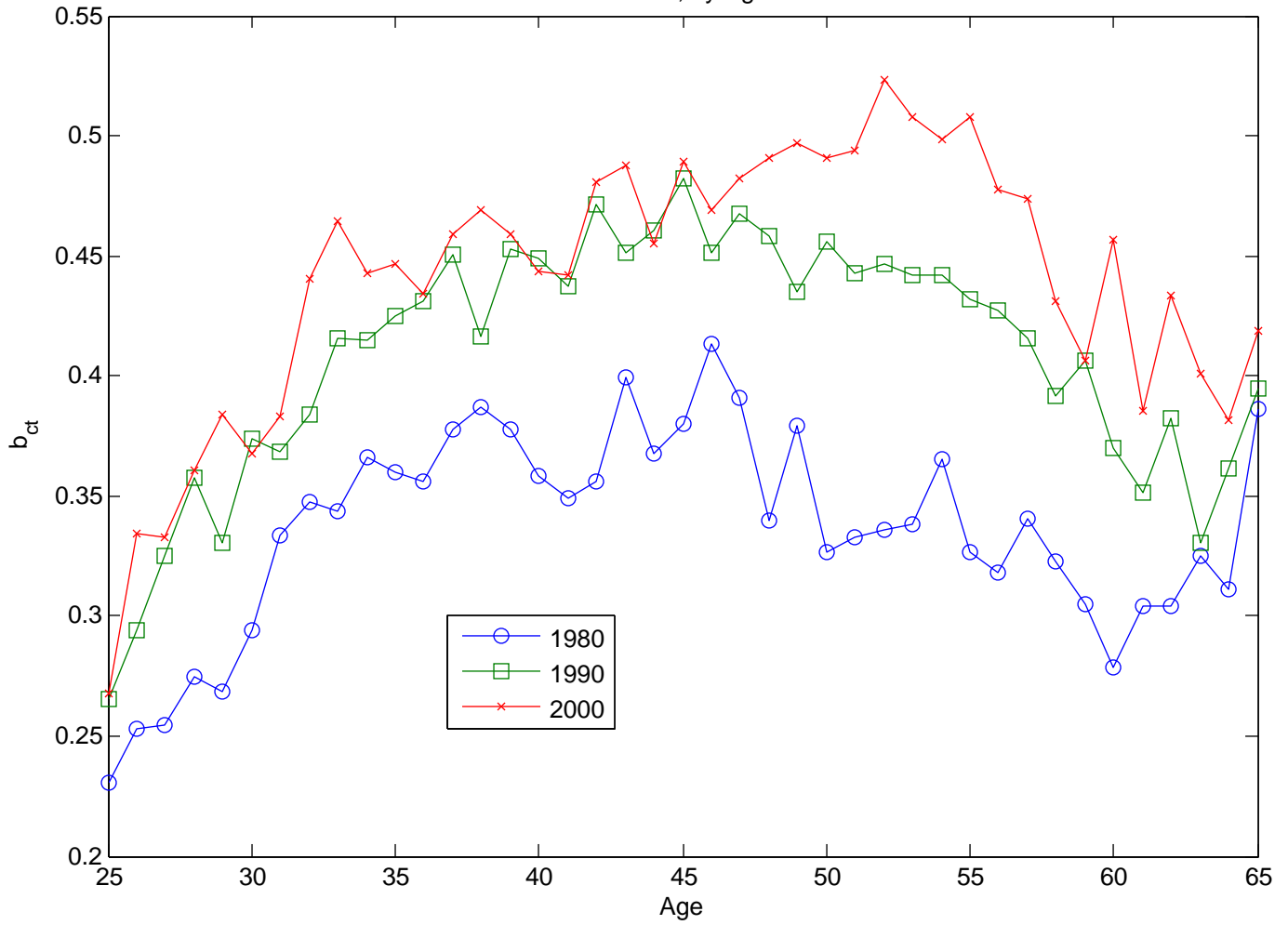


Figure 30: Mean High School Correlated Ability, μ_{act} , for $b_{ct} = .9b_{ct}^{OLS}$, 1940–1970, by Age = $c - t$

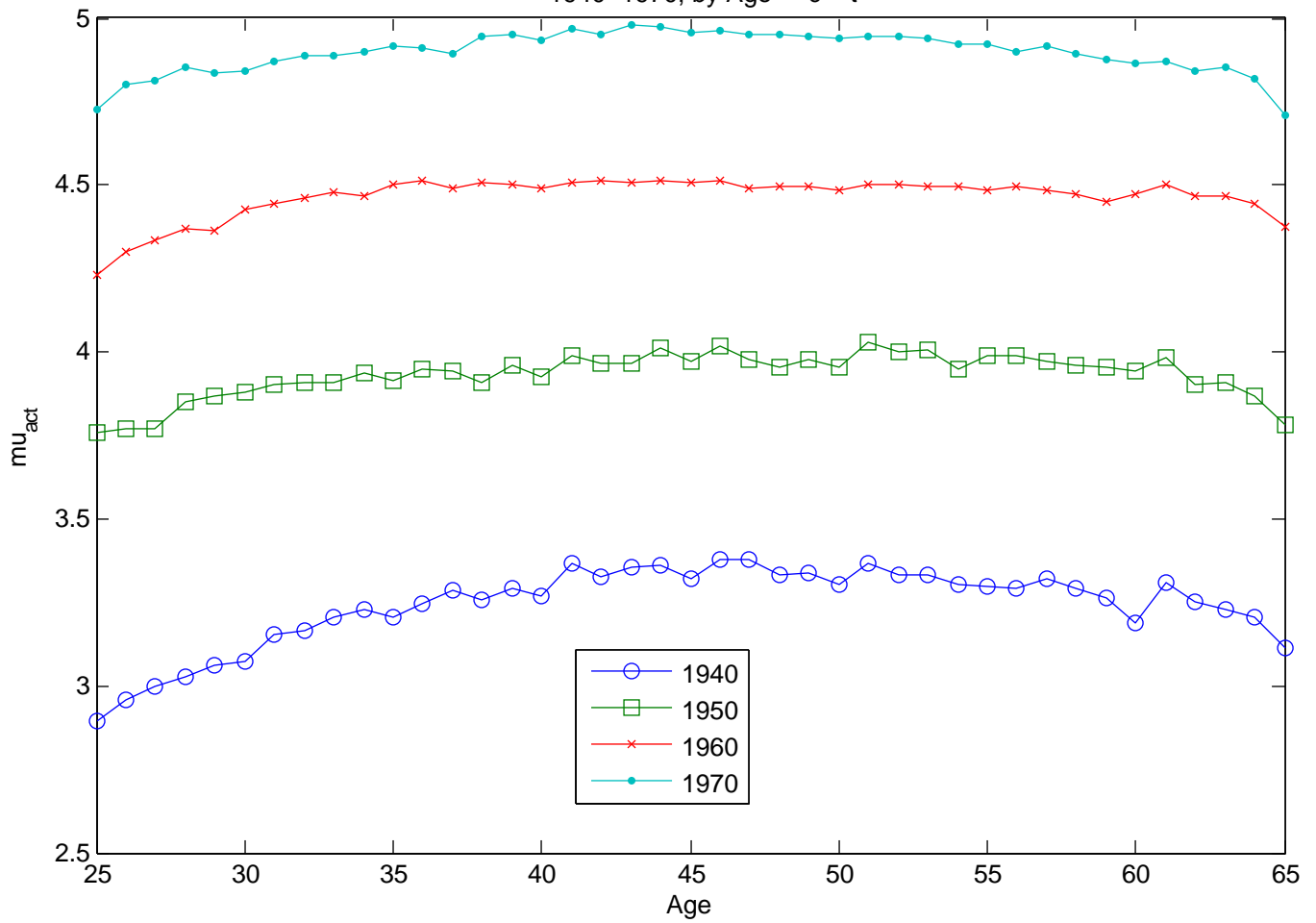


Figure 31: Mean High School Correlated Ability, μ_{act} , for $b_{ct} = .9b_{ct}^{OLS}$, 1980–2000, by Age = $c - t$

