"Although it is obvious that people acquire useful skills and knowledge, it is not obvious that these skills and knowledge are a form of capital, that this capital is in substantial part a product of deliberate investment, that it has grown in Western societies at a much faster rate than conventional (nonhuman) capital, and that its growth may well be the most distinctive feature of the economic system. It has been widely observed that increases in national output have been large compared with the increases of land, man-hours and physical reproducible capital. Investment in hc is probably the major explanation for this difference."

(T.W.Schultz - Presidential Address to the American Economics Association, 1961)
Among the first steps researchers took to quantify the investment nature of human capital was to calculate an internal rate of return to years of education.
Among the first steps researchers took to quantify the investment nature of human capital was to calculate an internal rate of return to years of education.

The internal rate of return is the rate of discount that equates the net income stream with and without an investment.
Among the first steps researchers took to quantify the investment nature of human capital was to calculate an internal rate of return to years of education.

The internal rate of return is the rate of discount that equates the net income stream with and without an investment.

Data needed for calculation

- Lifetime earnings stream for different levels of schooling.
- Direct costs of schooling
The early literature used cross-section data on earnings, age and years of schooling, which were more widely available from the Census.
The early literature used cross-section data on earnings, age and years of schooling, which were more widely available from the Census.

To calculate an internal rate of return from this data, researchers made the following assumptions:

1. the only cost of schooling is the foregone earnings;
The early literature used cross-section data on earnings, age and years of schooling, which were more widely available from the Census.

To calculate an internal rate of return from this data, researchers made the following assumptions:

1. the only cost of schooling is the foregone earnings;
2. labor market entry was immediate upon leaving school;
The early literature used cross-section data on earnings, age and years of schooling, which were more widely available from the Census.

To calculate an internal rate of return from this data, researchers made the following assumptions:

1. the only cost of schooling is the foregone earnings;
2. labor market entry was immediate upon leaving school;
3. working life is $n$ years, independent of years of schooling;
The early literature used cross-section data on earnings, age and years of schooling, which were more widely available from the Census.

To calculate an internal rate of return from this data, researchers made the following assumptions:

1. the only cost of schooling is the foregone earnings;
2. labor market entry was immediate upon leaving school;
3. working life is n years, independent of years of schooling;
4. there is no productivity growth in earnings, i.e., cross-section earnings = life cycle earnings;
The present value of lifetime earnings of an individual with $s$ years of schooling is

$$V(s, r) = \int_{t=s}^{n+s} y(s, t)e^{-rt} dt + (\int_{t=0}^{s} 0 \cdot e^{-rt} dt)$$

where the second term is foregone earnings from $t=0$ to $s$. 
The present value of lifetime earnings of an individual with \( s \) years of schooling is

\[
V(s, r) = \int_{t=s}^{n+s} y(s, t)e^{-rt}dt + (\int_{t=0}^{s} 0 \cdot e^{-rt}dt)
\]

where the second term is foregone earnings from \( t=0 \) to \( s \).

Denoting

\[
\rho(s, s + d) \equiv V(s, \rho) = V(s + d, \rho)
\]

as the internal rate of return to an additional \( d \) years of schooling, then \( \rho(s, s + d) \) is the solution to

\[
\int_{t=s}^{n+s} y(s, t)e^{-\rho t}dt = \int_{t=s+d}^{n+s+d} y(s + d, t)e^{-\rho t}dt
\]
In practice, the internal rate of return is usually unique because age-earnings profiles of two schooling groups only cross once when \( y(s,t) \) is approximated by a smooth function.

<table>
<thead>
<tr>
<th>Schooling Level</th>
<th>5-7</th>
<th>8</th>
<th>9-11</th>
<th>12</th>
<th>13-15</th>
<th>16</th>
<th>17+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whites/North</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whites/South</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Psacharopolous (1981) provided estimates by region and country type:

<table>
<thead>
<tr>
<th>Region or Country Type</th>
<th>N</th>
<th>Primary</th>
<th>Secondary</th>
<th>Higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>9</td>
<td>29</td>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>Asia</td>
<td>8</td>
<td>32</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Latin America</td>
<td>5</td>
<td>24</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>LDC Average</td>
<td>22</td>
<td>29</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>Intermediate</td>
<td>8</td>
<td>20</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>Advanced</td>
<td>14</td>
<td>-</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>
These early researchers concluded that the empirical estimates of the internal rate of return to years of schooling provided powerful evidence for the basic human capital hypothesis that views schooling as an investment which must be compensated with higher lifetime earnings.
Mincer (1958): compensating differences explain why persons with different levels of schooling receive different earnings over their lifetimes.
Mincer (1958): compensating differences explain why persons with different levels of schooling receive different earnings over their lifetimes.

The model makes the following assumptions:

(i) Individuals have identical abilities and opportunities, credit markets are perfect, earnings streams are certain.
Mincer (1958): compensating differences explain why persons with different levels of schooling receive different earnings over their lifetimes.

The model makes the following assumptions:

(i) Individuals have identical abilities and opportunities, credit markets are perfect, earnings streams are certain.

(ii) Individuals forego earnings while in school, but incur no direct costs and there are no consumption benefits from schooling.
Let $w(s) = \text{annual earnings with } s \text{ years of schooling (constant over the lifetime)}. \text{ Then the present value of lifetime earnings is}

$$V(s) = w(s) \int_s^{T+s} e^{-rt} dt = \frac{w(s)}{r} e^{-rs} (1 - e^{-rT})$$
Let $w(s) = \text{annual earnings with } s \text{ years of schooling (constant over the lifetime)}. \text{ Then the present value of lifetime earnings is}

$$V(s) = w(s) \int_s^{T+s} e^{-rt} \, dt = \frac{w(s)}{r} e^{-rs} (1 - e^{-rT})$$

Equating earnings streams across schooling levels $s$ and 0,

$$\frac{w(s)}{r} e^{-rs} (1 - e^{-rT}) = \frac{w(0)}{r} (1 - e^{-rT})$$
Let $w(s)$ = annual earnings with $s$ years of schooling (constant over the lifetime). Then the present value of lifetime earnings is

$$V(s) = w(s) \int_{s}^{T+s} e^{-rt} dt = \frac{w(s)}{r} e^{-rs} (1 - e^{-rT})$$

Equating earnings streams across schooling levels $s$ and 0,

$$\frac{w(s)}{r} e^{-rs} (1 - e^{-rT}) = \frac{w(0)}{r} (1 - e^{-rT})$$

$$w(s)e^{-rs} = w(0) \Rightarrow$$
Let $w(s) =$ annual earnings with $s$ years of schooling (constant over the lifetime). Then the present value of lifetime earnings is

$$V(s) = w(s) \int_s^{T+s} e^{-rt} dt = \frac{w(s)}{r} e^{-rs} (1 - e^{-rT})$$

Equating earnings streams across schooling levels $s$ and 0,

$$\frac{w(s)}{r} e^{-rs} (1 - e^{-rT}) = \frac{w(0)}{r} (1 - e^{-rT})$$

$$w(s) e^{-rs} = w(0) \implies$$

$$\log w(s) = \log w(0) + rs.$$
\[ \log w(s) = \log w(0) + rs. \]

The percentage increase in lifetime earnings associated with an additional year of schooling equals \( r \). By definition, the internal rate of return for an additional year must also equal \( r \), independent of \( s \). A regression of log-earnings on an intercept and years of schooling therefore yields an estimate of the internal rate of return, \( \rho_s = r \).
The Mincer schooling model does not explain the shape of the earnings profile after completing schooling, notably the rise in earnings with labor force experience.
The Mincer schooling model does not explain the shape of the earnings profile after completing schooling, notably the rise in earnings with labor force experience.

Becker (1965) developed an accounting framework for human capital investment on the job.
Let $E_t$ be potential earnings at age $t$, and express costs of investments in training $C_t$ as a fraction $k_t$ of potential earnings, $C_t = k_t E_t$. Let $\rho_t$ be the average return to training investments made at age $t$. Potential earnings at $t$ are

$$E_t = E_{t-1}(1 + \rho_{t-1}k_{t-1}) = \prod_{j=0}^{t-1}(1 + \rho_j k_j)E_0.$$
Attending school is defined as a period of full-time investment \((k_t = 1)\), which is assumed to take place at the beginning of life and to yield a rate of return \(\rho_s\) that is constant across all years of schooling. Assuming that the rate of return to post-school investment is constant over ages and equals \(\rho_0\), we can write

\[
\log E_t = \log E_0 + s \log(1 + \rho_s) + \sum_{j=s}^{t-1} \log(1 + \rho_0 k_j)
\]

\[
\approx \log E_0 + \rho_s s + \rho_0 \sum_{j=s}^{t-1} k_j,
\]

where the last approximation is made for small \(\rho_s\) and \(\rho_0\).
Mincer Earnings Function:

Mincer assumed a declining rate of post-school investment using a number of functional forms.

The one that motivates what is now called the Mincer earnings function is a linearly declining investment rate is

\[ k_{s+x} = \kappa \left( 1 - \frac{x}{T} \right) \]

where \( x = t - s \geq 0 \) is the amount of work experience as of age \( t \).
Under this form, the relationship between potential earnings, schooling and experience is given by:

\[ \log E_{x+s} \approx [\log E_0 - \kappa \rho_0] + \rho_s s + \left( \rho_0 \kappa + \frac{\rho_0 \kappa}{2T} \right) x - \frac{\rho_0 \kappa}{2T} x^2. \]
Under this form, the relationship between potential earnings, schooling and experience is given by:

\[ \log E_{x+s} \approx \left[ \log E_0 - \kappa \rho_0 \right] + \rho_s s + \left( \rho_0 \kappa + \frac{\rho_0 \kappa}{2T} \right)x - \frac{\rho_0 \kappa}{2T} x^2. \]

Observed earnings are potential earnings less investment costs, producing the relationship for observed earnings known as the Mincer equation:

\[ \log w(s,x) \approx \log E_{x+s} - \kappa \left( 1 - \frac{x}{T} \right) \]

\[ = [\log E_0 - \kappa \rho_0 - \kappa] + \rho_s s + \left( \rho_0 \kappa + \frac{\rho_0 \kappa}{2T} + \frac{\kappa}{T} \right)x - \frac{\rho_0 \kappa}{2T} x^2 \]

\[ = \beta_0 + \beta_1 s + \beta_2 x - \beta_3 x^2 \]
Log earnings are linear in years of schooling, and linear and quadratic in years of labor market experience.

In this formulation, $\rho_s$ is an average rate of return across all schooling investments and not, in general, an internal rate of return.

$$\log w(s,x) = \beta_0 + \beta_1 s + \beta_2 x - \beta_3 x^2$$
Mincer reports the following estimates of the human capital earnings function, estimated on white men in non-farm occupations using 1960 Census data, where t-ratios are shown in parentheses.
Mincer reports the following estimates of the human capital earnings function, estimated on white men in non-farm occupations using 1960 Census data, where t-ratios are shown in parentheses.

\[ \log y = 7.58 + 0.070s \quad (43.8) \quad R^2 = 0.067 \]
Mincer reports the following estimates of the human capital earnings function, estimated on white men in non-farm occupations using 1960 Census data, where t-ratios are shown in parentheses.

\[ \log y = 7.58 + 0.070 s \quad R^2 = 0.067 \]

\[ \log y = 6.20 + 0.107 s + 0.081 x - 0.0012 x^2 \quad R^2 = 0.285 \]
Mincer reports the following estimates of the human capital earnings function, estimated on white men in non-farm occupations using 1960 Census data, where t-ratios are shown in parentheses.

\[
\log y = 7.58 + 0.070s \quad R^2 = 0.067 \\
\begin{array}{c}
(43.8)
\end{array}
\]

\[
\log y = 6.20 + 0.107s + 0.081x - 0.0012x^2 \quad R^2 = 0.285 \\
\begin{array}{c}
(72.3) \\
(75.5) \\
(-55.8)
\end{array}
\]

\[
\log y = 4.87 + 0.255s - 0.0029s^2 - 0.0043xs + 0.148x - 0.0018x^2 \quad R^2 = 0.309 \\
\begin{array}{c}
(23.4) \\
(-7.1) \\
(-31.8) \\
(63.7) \\
(-66.2)
\end{array}
\]
Becker’s Woytinsky Lecture (1967), Rosen (1977)

What determines the level of schooling in the compensating difference model? In the accounting framework?
Becker’s Woytinsky Lecture

What determines the level of schooling in the compensating difference model? In the accounting framework?

Both models are silent about why people vary in their schooling choice.
We consider a simple supply-demand framework in which individuals may differ in their demand for human capital as well as in the supply of funds available for human capital investment.

Letting $A = \text{ability}$ and $y = \text{income}$,

$$\ln y = h(s, A)$$

where

$$h_A > 0, h_s > 0, h_{ss} < 0.$$
We consider a simple supply-demand framework in which individuals may differ in their demand for human capital as well as in the supply of funds available for human capital investment.

Letting $A = \text{ability}$ and $y = \text{income}$,

$$\ln y = h(s, A)$$

where

$$h_A > 0, h_s > 0, h_{ss} < 0.$$ 

Assume that the market rate of discount for a person with family background characteristic $Z$, reflecting credit market access, is

$$r = r(Z).$$
The earnings stream, over $n$ years, associated with completing $s$ years of schooling is

$$V(s) = \max_y y(s) \int_s^{n+s} e^{-rt} dt$$
The earnings stream, over $n$ years, associated with completing $s$ years of schooling is

$$V(s) = \max_s y(s) \int_s^{n+s} e^{-rt} dt$$

$$= \max_s \frac{a}{r} e^{-rs} e^{h(s-A)} \text{ with } a = 1 - e^{-rn}. $$
The earnings stream, over \( n \) years, associated with completing \( s \) years of schooling is

\[
V(s) = \max_s y(s) \int_s^{n+s} e^{-rt} dt
\]

\[
= \max_s \frac{a}{r} e^{-rs} e^{h(s,A)} \text{ with } a = 1 - e^{-rn}.
\]

Choosing \( s \) to maximize \( V(s) \):

\[
\frac{\partial V(s)}{\partial s} = \frac{a}{r} e^{-rs+h(s,A)} (h_s - r) = 0 \implies
\]
The earnings stream, over $n$ years, associated with completing $s$ years of schooling is

$$V(s) = \max_s y(s) \int_s^{n+s} e^{-rt} \, dt$$

$$= \max_s \frac{a}{r} e^{-rs} e^{h(s,A)} \text{ with } a = 1 - e^{-rn}.$$  

Choosing $s$ to maximize $V(s)$:

$$\frac{\partial V(s)}{\partial s} = \frac{a}{r} e^{-rs+h(s,A)} (h_s - r) = 0 \Rightarrow$$

$$h_s(s, A) = r(Z) \Rightarrow$$
The earnings stream, over $n$ years, associated with completing $s$ years of schooling is

$$ V(s) = \max_s y(s) \int_s^{n+s} e^{-rt} \, dt $$

$$ = \max_s \frac{a}{r} e^{-rs} e^{h(s,A)} \text{ with } a = 1 - e^{-rn}. $$

Choosing $s$ to maximize $V(s)$:

$$ \frac{\partial V(s)}{\partial s} = \frac{a}{r} e^{-rs+h(s,A)} (h_s - r) = 0 \Rightarrow $$

$$ h_s(s, A) = r(Z) \Rightarrow $$

$$ s = s(A, r(Z)) $$
\[ s = s(A, r(Z)) \]

If everyone has the same Z and A, then all would choose the same schooling level.

Taking the total derivative of the first order condition,

\[ \frac{ds}{dA} = \frac{-h_{sA}}{h_{ss}} > 0 \text{ if } h_{sA} > 0 \]

\[ \frac{ds}{dr} = \frac{1}{h_{ss}} < 0. \]
Becker distinguishes between two views of the world.

If everyone faced the same market interest rate $r$ (there was equality of opportunity), then schooling choices would differ only due to variation in ability or motivation. Those who invest more in schooling have a higher marginal return function, that is, they are better at it - the *elitist view*. 
Becker distinguishes between two views of the world.

If everyone faced the same market interest rate $r$ (there was equality of opportunity), then schooling choices would differ only due to variation in ability or motivation. Those who invest more in schooling have a higher marginal return function, that is, they are better at it - the *elitist view*.

On the other hand, if individuals are identical in ability, the egalitarian view, and the only source of differences across people is in the marginal cost of financing schooling investments, schooling differences would arise due to extrinsic factors.
Inequality in schooling outcomes may arise for both reasons.

Demand side schooling intervention

   Early childhood intervention that increased the child's human capital stock upon school entry
Inequality in schooling outcomes may arise for both reasons.

Demand side schooling intervention
   Early childhood intervention that increased the child's human capital stock upon school entry

Supply side schooling intervention
   Subsidized student loans
Ben-Porath (1967): Optimal Human Capital Accumulation

Unlike the compensating difference or the accounting identity framework, Ben-Porath develops a model of the optimal allocation of resources to the production of human capital.

The essential idea was that human capital is produced according to some technology that combines an individual's time with purchased goods.

The original formulation was in continuous time.
Let $H_t$ denote the (homogeneous) stock of human capital embodied in a person at age $t$, $R$ the rental price of a unit of human capital, $I_t$ the proportion of work time spent investing in human capital ($0 \leq I_t \leq 1$), $D_t$ the amount of purchased inputs used to produce human capital and $p_D$ their per-unit price. Human capital also depreciates at rate $\theta$. 
Let $H_t$ denote the (homogeneous) stock of human capital embodied in a person at age $t$, $R$ the rental price of a unit of human capital, $I_t$ the proportion of work time spent investing in human capital ($0 \leq I_t \leq 1$), $D_t$ the amount of purchased inputs used to produce human capital and $p_D$ their per-unit price. Human capital also depreciates at rate $\theta$.

The production function for human capital is

$$\dot{H}_t = F(I_t, H_t, D_t; \alpha) - \theta H_t.$$  

$H(0)$ is the human capital endowment and $\alpha$ is “ability” to produce human capital.
Earnings, net of the cost of purchases inputs, are given by

\[ w_t = RH_t (1 - I_t) - p_D D_t, \]

where workers are paid for time not spent investing in human capital.
Earnings, net of the cost of purchases inputs, are given by

$$w_t = RH_t(1 - I_t) - p_D D_t,$$

where workers are paid for time not spent investing in human capital.

The individual maximizes the present value of net earnings

$$V(H_0) = \max_{I_t, D_t} \int_0^T e^{-rt} [RH_t(1 - I_t) - p_D D_t] dt$$

given the human capital production function and endowments.
Discrete Time – no purchased inputs:

Production Function:

\[ H_t = G(H_{t-1}, I_{t-1}; \alpha) \]

where

\[ G_I > 0, G_H > 0, G_{II} < 0 \]
\[ G_\alpha > 0, G_{I\alpha} > 0, G_{IH} > 0 \]
Discrete Time – no purchased inputs:

Production Function:

\[ H_t = G(H_{t-1}, I_{t-1}; \alpha) \]

where

\[ G_I > 0, \ G_H > 0, \ G_{II} < 0 \]
\[ G_\alpha > 0, \ G_{I\alpha} > 0, \ G_{IH} > 0 \]

Earnings:

\[ Y_t = RH_t(1 - I_t) \]
Bellman equation:

\[ V_t(H_t) = \max_{\{I_t\}} RH_t(1 - I_t) + \delta V_{t+1}(H_{t+1}) \]
Bellman equation:

\[ V_t(H_t) = \max_{\{I_t\}} RH_t(1 - I_t) + \delta V_{t+1}(H_{t+1}) \]

with

\[ H_{t+1} = G(H_t, I_t; \alpha) \]

\[ 0 \leq I_t \leq 1 \text{ for all } t = 1, \ldots, T \]

\[ V_{T+1} = 0 \]
Period T:

\[ V_T(H_T) = \max_{I_T} R H_T (1 - I_T) \]

\[ \Rightarrow I_T^* = 0 \]
Period T:

\[ V_T(H_T) = \max_{I_T} RH_T(1 - I_T) \]

\[ \Rightarrow I_T^* = 0 \]

Thus,

\[ V_T(H_T) = RH_T \]

\[ = RG(H_{T-1}, I_{T-1}; \alpha) \]
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RH_T \]

\[ = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RH_T \]

\[ = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]

\[ \frac{\partial V_{T-1}}{\partial I_{T-1}} = -RH_{T-1} + \delta RG_I \geq 0 \]
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RH_{T} \]

\[ = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]

\[ \frac{\partial V_{T-1}}{\partial I_{T-1}} = -RH_{T-1} + \delta RG_{I} \geq 0 \]

Investment Decision Rule:
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RH_T \]

\[ = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]

\[ \frac{\partial V_{T-1}}{\partial I_{T-1}} = -RH_{T-1} + \delta RG_I \geq 0 \]

Investment Decision Rule:

\[ I_{T-1}^* = 0 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, 0; \alpha) < 0 \]
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} \ RH_{T-1}(1 - I_{T-1}) + \delta RH_T \]

\[ = \max_{I_{T-1}} \ RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]

\[ \frac{\partial V_{T-1}}{\partial I_{T-1}} = -RH_{T-1} + \delta RG_I \geq 0 \]

Investment Decision Rule:

\[ I^*_{T-1} = 0 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, 0; \alpha) < 0 \]

\[ 0 < I^*_{T-1} < 1 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, I^*_{T-1}; \alpha) = 0 \]
Period T-1:

\[ V_{T-1}(H_{T-1}) = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RH_T \]
\[ = \max_{I_{T-1}} RH_{T-1}(1 - I_{T-1}) + \delta RG(H_{T-1}, I_{T-1}) \]

\[ \frac{\partial V_{T-1}}{\partial I_{T-1}} = -RH_{T-1} + \delta RG_I \geq 0 \]

Investment Decision Rule:

\[ I_{T-1}^* = 0 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, 0; \alpha) < 0 \]

\[ 0 < I_{T-1}^* < 1 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, I_{T-1}^*; \alpha) = 0 \]

\[ I_{T-1}^* = 1 \text{ if } -H_{T-1} + \delta G_I(H_{T-1}, 1; \alpha) > 0 \]
If there is an interior solution, i.e., $0 < I_{T-1} < 1$, then

$$H_{T-1} = \delta G_I(H_{T-1}, I_{T-1}; \alpha)$$

$$\Rightarrow \quad I_{T-1} = I_{T-1}^*(H_{T-1}, \delta; \alpha)$$
If there is an interior solution, i.e., $0 < I_{T-1} < 1$, then

$$H_{T-1} = \delta G_I(H_{T-1}, I_{T-1}; \alpha)$$

$$\Rightarrow \quad I_{T-1} = I^*_T(H_{T-1}, \delta; \alpha)$$

Note that the optimal investment time is independent of $R$, the human capital rental price.

Why?
If there is an interior solution, i.e., \(0 < I_{T-1} < 1\), then

\[
H_{T-1} = \delta G_I(H_{T-1}, I_{T-1}; \alpha)
\]

\[
\Rightarrow \quad I_{T-1} = I_{T-1}^*(H_{T-1}, \delta; \alpha)
\]

Note that the optimal investment time is independent of \(R\), the human capital rental price.

Why?
If there is an interior solution, i.e., $0 < I_{T-1} < 1$, then

$$H_{T-1} = \delta G_I(H_{T-1}, I_{T-1}; \alpha)$$

$$\Rightarrow \quad I_{T-1} = I^*_{T-1}(H_{T-1}, \delta; \alpha)$$

Note that the optimal investment time is independent of $R$, the human capital rental price

because the only cost of investing is foregone earnings.
Totally differentiating the first-order condition for the case of an interior solution, implies the following comparative static results:
Totally differentiating the first-order condition for the case of an interior solution, implies the following comparative static results:

\[
\frac{dI_{T-1}}{dH_{T-1}} = \frac{1 - \delta G_{IH}}{\delta G_{II}} \geq 0 \quad \text{as} \quad G_{IH} \geq \frac{1}{\delta}
\]
Totally differentiating the first-order condition for the case of an interior solution, implies the following comparative static results:

\[
\frac{dI_{T-1}}{dH_{T-1}} = \frac{1 - \delta G_{IH}}{\delta G_{II}} \geq 0 \quad \text{as} \quad G_{IH} \geq \frac{1}{\delta}
\]

\[
\frac{dI_{T-1}}{d\delta} = \frac{-G_I}{\delta G_{II}} > 0,
\]
Totally differentiating the first-order condition for the case of an interior solution, implies the following comparative static results:

\[
\frac{dI_{T-1}}{dH_{T-1}} = \frac{1 - \delta G_{IH}}{\delta G_{II}} \geq 0 \quad \text{as} \quad G_{IH} \geq \frac{1}{\delta}
\]

\[
\frac{dI_{T-1}}{d\delta} = \frac{-G_I}{\delta G_{II}} > 0,
\]

\[
\frac{dI_{T-1}}{d\alpha} = \frac{-G_{I\alpha}}{G_{II}} > 0
\]
To get further, consider the following functional form for the human capital production function:

\[ H_{t+1} = \alpha (H_t I_t) \beta + (1 - \theta) H_t, \]

where \( \beta < 1 \).
To get further, consider the following functional form for the human capital production function:

\[ H_{t+1} = \alpha (H_t I_t)^\beta + (1 - \theta) H_t, \]

where \( \beta < 1 \). Then,

\[ G_H = \alpha \beta H_t^{\beta - 1} I_t^B + (1 - \theta) > 0 \]
To get further, consider the following functional form for the human capital production function:

\[ H_{t+1} = \alpha (H_t I_t)^\beta + (1 - \theta) H_t, \]

where \( \beta < 1 \). Then,

\[ G_H = \alpha \beta H_t^{\beta-1} I_t^\beta + (1 - \theta) > 0 \]

\[ G_I = \alpha \beta H_t^\beta I_t^{\beta-1} > 0 \]
To get further, consider the following functional form for the human capital production function:

\[ H_{t+1} = \alpha (H_t I_t)^\beta + (1 - \theta) H_t, \]

where \( \beta < 1 \). Then,

\[
G_H = \alpha \beta H_t^{\beta-1} I_t^\beta + (1 - \theta) > 0
\]

\[
G_I = \alpha \beta H_t^{\beta} I_t^{\beta-1} > 0
\]

\[
G_{HI} = \alpha \beta^2 (H_t I_t)^{\beta-1} > 0
\]
To get further, consider the following functional form for the human capital production function:

$$H_{t+1} = \alpha (H_t I_t) + (1 - \theta) H_t,$$

where $\beta < 1$. Then,

$$G_H = \alpha \beta H_t^{\beta-1} I_t^\beta + (1 - \theta) > 0$$

$$G_I = \alpha \beta H_t^\beta I_t^{\beta-1} > 0$$

$$G_{HI} = \alpha \beta^2 (H_t I_t)^{\beta-1} > 0$$

Note that because $G_I \to \infty$ as $I \to 0$, $I_t^* \neq 0$ for all $t < T$. $I_T^*$ is still zero because there is no future return.
T-1:

\[
\frac{\partial V_{T-1}}{\partial I_{T-1}}|_{I_{T-1}=1} = -RH_{T-1} + R\delta \alpha \beta H^\beta_{T-1} > 0 \text{ as }
\]

\[
H_{T-1} < (\delta \alpha \beta)^{\frac{1}{1-\beta}} = H^*_{T-1}
\]
T-1:

$$\frac{\partial V_{T-1}}{\partial I_{T-1}}|_{I_{T-1}=1} = -RH_{T-1} + R\delta \alpha \beta H_{T-1}^{\beta} > 0$$

as

$$H_{T-1} < (\delta \alpha \beta)^{\frac{1}{1-\beta}} = H_{T-1}^*$$

Thus, $$I_{T-1}^* = 1$$ if the stock of human capital entering period T-1 is smaller than $$H_{T-1}^*$$.
If the stock exceeds that threshold value, then the investment is interior:

\[- H_{T-1} + \delta \alpha \beta H_{T-1} \beta I_{T-1}^{\beta-1} = 0,\]
If the stock exceeds that threshold value, then the investment is interior:

\[- H_{T-1} + \delta \alpha \beta H_{T-1}^{1-\beta} I_{T-1}^{\beta-1} = 0,\]

which yields

\[I_{T-1}^* = (\delta \alpha \beta)^{1-\beta} H_{T-1}^{-1}\]
If the stock exceeds that threshold value, then the investment is interior:

\[- H_{T-1} + \delta \alpha \beta H_T^\beta I_{T-1}^{\beta-1} = 0,\]

which yields

\[I_{T-1}^* = (\delta \alpha \beta)^{\frac{1}{1-\beta}} H_{T-1}^{-1}\]

and

\[H_{T-1} I_{T-1}^* = (\delta \alpha \beta)^{\frac{1}{1-\beta}}\]
If the stock exceeds that threshold value, then the investment is interior:

\[- H_{T-1} + \delta \alpha \beta H_{T-1}^\beta I_{T-1}^{\beta-1} = 0,\]

which yields

\[I_{T-1}^* = (\delta \alpha \beta)^{\frac{1}{1-\beta}} H_{T-1}^{-1}\]

and

\[H_{T-1} I_{T-1}^* = (\delta \alpha \beta)^{\frac{1}{1-\beta}}\]

Thus, investment declines with the stock, though never to zero.
T-2:

\[ V_{T-2}(H_{T-2}) = \max_{I_{T-2}} \{ R_{H_{T-2}}(1 - I_{T-2}) + \delta V_{T-1}(H_{T-1}|H_{T-2}) \} \]
T-2:

\[ V_{T-2}(H_{T-2}) = \max_{I_{T-2}} \{ RH_{T-2}(1 - I_{T-2}) + \delta V_{T-1}(H_{T-1}|H_{T-2}) \} \]

\[ = \max_{I_{T-2}} \{ RH_{T-2}(1 - I_{T-2}) + \]
T-2:

\[ V_{T-2}(H_{T-2}) = \]
\[
\max_{I_{T-2}} \{ RH_{T-2}(1 - I_{T-2}) + \delta V_{T-1}(H_{T-1}|H_{T-2}) \}
\]
\[
= \max_{I_{T-2}} \{ RH_{T-2}(1 - I_{T-2}) + \delta \max(V_{T-1}(H_{T-1}; I_{T-1}^* = 1|H_{T-2}), V_{T-1}(H_{T-1}; I_{T-1}^* < 1|H_{T-2}) \} \]
The following two propositions hold. They are left to you to prove.
The following two propositions hold. They are left to you to prove.

1. $H_{T-2}I_{T-2}^* > H_{T-1}I_{T-1}^*$: it is optimal to produce more human capital at T-2 than at T-1. More generally, the amount of human capital produced declines with age.
The following two propositions hold. They are left to you to prove.

1. $H_{T-2}I_{T-2}^* > H_{T-1}I_{T-1}^*$: it is optimal to produce more human capital at T-2 than at T-1. More generally, the amount of human capital produced declines with age.

2. If $I_{T-1}^* = 1$, the $I_{T-2}^* = 1$: once one stops producing human capital, one never specializes in the production of human capital again.
The following two propositions hold. They are left to you to prove.

1. $H_{T-2}I_{T-2}^* > H_{T-1}I_{T-1}^*$: it is optimal to produce more human capital at T-2 than at T-1. More generally, the amount of human capital produced declines with age.

2. If $I_{T-1}^* = 1$, the $I_{T-2}^* = 1$: once one stops producing human capital, one never specializes in the production of human capital again.

What is the real world analog to specialization in the production of human capital?
The Ben-Porath (1967) – Griliches (1977) Wage Function

How should we interpret a wage (earnings) function?
The Ben-Porath (1967) – Griliches (1977) Wage Function

How should we interpret a wage (earnings) function?

In the Ben-Porath model, the wage at any age is determined from a pricing equation for homogeneous human capital, namely as the (competitively determined) rental price per unit of human capital times the stock of human capital:

$$w_a = rH_a$$
It is useful to put this pricing equation into a market setting. Suppose that the economy can be described by an aggregate production function that depends on the total human capital supplied to the economy $\tilde{H} = \sum \sum H_{ia}$ and the amount of physical capital:

$$Y = F(\tilde{H}, K)$$

where there are diminishing returns in both inputs.
It is useful to put this pricing equation into a market setting. Suppose that the economy can be described by an aggregate production function that depends on the total human capital supplied to the economy $\tilde{H} = \sum \sum H_{ia}$ and the amount of physical capital:

$$Y = F(\tilde{H}, K)$$

where there are diminishing returns in both inputs.

The rental price per unit of human capital is the MP of the aggregate human capital used in production,

$$r(\tilde{H}, K) = \partial F / \partial \tilde{H}$$
Taking logs of the pricing equation,

$$\log w_a = \log(r) + \log H_a.$$ 

In general, the rental price will vary with calendar time as there are aggregate shocks to production or preferences.
Taking logs of the pricing equation,

\[ \log w_a = \log(r) + \log H_a. \]

In general, the rental price will vary with calendar time as there are aggregate shocks to production or preferences.

A wage function consistent with Mincer's formulation is easily derived from the competitive skill market model. If the skill production function takes the form,

\[ H_a = H_0 \exp\{\beta_1 s + \beta_2 x - \beta_3 x^2 + \epsilon\}, \]
Taking logs of the pricing equation,

$$\log w_a = \log(r) + \log H_a.$$ 

In general, the rental price will vary with calendar time as there are aggregate shocks to production or preferences.

A wage function consistent with Mincer's formulation is easily derived from the competitive skill market model. If the skill production function takes the form,

$$H_a = H_0 \exp\{\beta_1 s + \beta_2 x - \beta_3 x^2 + \epsilon\},$$

and the market wage offer function is

$$\log w_a = \log(r) + \log H_0 + \beta_1 s + \beta_2 x - \beta_3 x^2 + \epsilon.$$
This rationale for this wage function was first noted by Grilliches (1977, Presidential Address in 1975).
This rationale for this wage function was first noted by Grilliches (1977, Presidential Address in 1975).

The constant term in the Mincer formulation can be interpreted as a composite of

(i) the skill rental price (which may depend on calendar time) and
This rationale for this wage function was first noted by Grilliches (1977, Presidential Address in 1975).

The constant term in the Mincer formulation can be interpreted as a composite of

(i) the skill rental price (which may depend on calendar time) and

(ii) the level of the individual's (non-school) pre-market skills, which may differ among individuals either because of genetic endowment or family investment behavior.
This rationale for this wage function was first noted by Grilliches (1977, Presidential Address in 1975).

The constant term in the Mincer formulation can be interpreted as a composite of

(i) the skill rental price (which may depend on calendar time) and

(ii) the level of the individual's (non-school) pre-market skills, which may differ among individuals either because of genetic endowment or family investment behavior.

Shocks to wage offers, other than through aggregate production shocks incorporated in the rental price, reflect idiosyncratic shocks to skills (e.g., random mental or physical lapses or enhancements).
In this framework, the effect of school attainment on the wage is fixed by the technology of the educational system in producing market valued skills and does not have the interpretation of an internal rate of return.

Likewise, the effect of work experience on the wage is fixed by the technology of skill acquisition on the job, i.e., by learning by doing.
Ability Bias:
In the Ben-Porath and Becker models of human capital accumulation, the stock of human capital depended on an individual's ability to produce human capital.
Ability Bias:

In the Ben-Porath and Becker models of human capital accumulation, the stock of human capital depended on an individual's ability to produce human capital.

Adopting the Ben-Porath/Griliches interpretation of the wage function as a human capital pricing equation (omitting work experience for convenience),

\[
\log y_a = \log r + \log H_a(s, A)
\]

\[
H_a = \exp(\beta_0 + \beta_s s_a + \beta_A A + u)
\]
Ability Bias:

In the Ben-Porath and Becker models of human capital accumulation, the stock of human capital depended on an individual's ability to produce human capital.

Adopting the Ben-Porath/Griliches interpretation of the wage function as a human capital pricing equation (omitting work experience for convenience),

\[
\log y_a = \log r + \log H_a(s, A)
\]

\[
H_a = \exp(\beta_0 + \beta_s s_a + \beta_A A + u)
\]

What is \( \beta_0 \) in the Ben-Porath - Griliches framework?
Ability Bias:

\[ \log y_a = \log r + \beta_0 + \beta_s s_a + \beta_A A + u \]

- \( y_a \) = wage paid to an individual of age \( a \)
- \( r \) = per unit human capital rental price
- \( s_a \) = schooling level
- \( A \) = “ability”
Ability Bias:

\[
\log y_a = \log r + \beta_0 + \beta_s s_a + \beta_A A + u
\]

- \(y_a\) = wage paid to an individual of age \(a\)
- \(r\) = per unit human capital rental price
- \(s_a\) = schooling level
- \(A\) = “ability”

Is this formulation consistent with the role of ability in the Ben-Porath model?
Suppose that ability is unobserved and that we estimate

\[ \log y = a + bs + e. \]
Suppose that ability is unobserved and that we estimate

\[ \log y = a + bs + e. \]

The ols estimate is

\[ \hat{b} = \frac{\text{cov}(s, \log y)}{\text{var}(s)}. \]
Suppose that ability is unobserved and that we estimate

$$\log y = a + bs + e.$$ 

The ols estimate is

$$\hat{b} = \frac{\text{cov}(s, \log y)}{\text{var}(s)} = \frac{\text{cov}(s, a + \beta_s s + \beta_A A + u)}{\text{var}(s)}.$$
Suppose that ability is unobserved and that we estimate

\[
\log y = a + bs + e.
\]

The ols estimate is

\[
\hat{b} = \frac{\text{cov}(s, \log y)}{\text{var}(s)}
\]

\[
= \frac{\text{cov}(s, a + \beta_s s + \beta_A A + u)}{\text{var}(s)}
\]

\[
= \beta_s + \beta_A \frac{\text{cov}(s, A)}{\text{var}(s)}
\]
Suppose that ability is unobserved and that we estimate

$$ \log y = a + bs + e. $$

The OLS estimate is

$$ \hat{b} = \frac{\text{cov}(s, \log y)}{\text{var}(s)} $$

$$ = \frac{\text{cov}(s, a + \beta_s s + \beta_A A + u)}{\text{var}(s)} $$

$$ = \beta_s + \beta_A \frac{\text{cov}(s, A)}{\text{var}(s)} $$

$$ = \beta_s + \beta_A \beta_{As} $$
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

  a) $\beta_A$ - the size of the effect of ability on earnings (conditional on schooling),
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

a) $\beta_A$ - the size of the effect of ability on earnings (conditional on schooling),

b) $\beta_{As}$ - the size of the regression coefficient of ability on schooling.
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

a) $\beta_A$ - the size of the effect of ability on earnings (conditional on schooling),

b) $\beta_{As}$ - the size of the regression coefficient of ability on schooling.

In what ways can we eliminate the bias?
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

a) $\beta_A$ - the size of the effect of ability on earnings (conditional on schooling),

b) $\beta_{As}$ - the size of the regression coefficient of ability on schooling.

In what ways can we eliminate the bias?

1. Measure ability
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

a) \( \beta_A \) - the size of the effect of ability on earnings (conditional on schooling),

b) \( \beta_{As} \) - the size of the regression coefficient of ability on schooling.

In what ways can we eliminate the bias?

1. Measure ability
2. Use a within (or fixed-effects) estimator
The ols estimate is biased if schooling and ability are correlated. The size of the bias depends on two things:

   a) $\beta_A$ - the size of the effect of ability on earnings (conditional on schooling),
   b) $\beta_{As}$ - the size of the regression coefficient of ability on schooling.

In what ways can we eliminate the bias?

   1. Measure ability
   2. Use a within (or fixed-effects) estimator
   3. Instrumental variables – find a variable correlated with schooling but not with ability
Measuring Ability

A number of studies have attempted to eliminate the ability bias by using test scores as measures of ability.

There are two problems that researchers encounter with this approach:
Measuring Ability

A number of studies have attempted to eliminate the ability bias by using test scores as measures of ability.

There are two problems that researchers encounter with this approach:

1. test scores are imperfect measures of ability
Measuring Ability

A number of studies have attempted to eliminate the ability bias by using test scores as measures of ability.

There are two problems that researchers encounter with this approach:

1. test scores are imperfect measures of ability

2. test scores are often measured subsequent to some individuals having completed their schooling – thus, test scores may themselves depend on schooling.
Griliches and Mason (1972)

Data:

Sample of 3000 WWII veterans, all of whom took an “ability” test (the Armed Forces Qualifying Test or AFQT) upon entry into the military.
Griliches and Mason (1972)

Data:

Sample of 3000 WWII veterans, all of whom took an “ability” test (the Armed Forces Qualifying Test or AFQT) upon entry into the military.

Interviewed as a supplemental sample in the 1964 CPS.
Griliches and Mason (1972)

Data:

Sample of 3000 WWII veterans, all of whom took an “ability” test (the Armed Forces Qualifying Test or AFQT) upon entry into the military.

Interviewed as a supplemental sample in the 1964 CPS.

GM chose 1454 respondents who were not in school and who were employed full-time at the survey date.
Griliches and Mason (1972)

Data:

Sample of 3000 WWII veterans, all of whom took an “ability” test (the Armed Forces Qualifying Test or AFQT) upon entry into the military.

Interviewed as a supplemental sample in the 1964 CPS.

GM chose 1454 respondents who were not in school and who were employed full-time at the survey date.

Many veterans used the GI bill, which subsidized college tuition and provided limited stipends to attend college.
For those who returned to school, the ability test could not have been affected by the amount of schooling they subsequently received,
For those who returned to school, the ability test could not have been affected by the amount of schooling they subsequently received, although it could have been affected by the amount of schooling they had completed prior to entering the military.
For those who returned to school, the ability test could not have been affected by the amount of schooling they subsequently received, although it could have been affected by the amount of schooling they had completed prior to entering the military.

GM estimate an earnings function that separately accounts for the amount of schooling received prior to entering the military (SB) and the incremental amount of schooling received subsequent to leaving the military (SI).
They report four specifications, with and without the AFQT variable and with and without additional variables.

<table>
<thead>
<tr>
<th></th>
<th>Schooling Before</th>
<th>Schooling Increment</th>
<th>includes AFQT</th>
<th>Additional variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.0502</td>
<td>.0528</td>
<td>NO</td>
<td>Age, Length of Service</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.0418</td>
<td>.0475</td>
<td>YES</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.0379</td>
<td>.0496</td>
<td>NO</td>
<td>+Father’s education and occupation, Region</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.007)</td>
<td></td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>.0328</td>
<td>.0462</td>
<td>YES</td>
<td>&quot;</td>
</tr>
<tr>
<td></td>
<td>(.005)</td>
<td>(.007)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary of Results:

The inclusion of AFQT reduces the effect of SB by 17%, but reduces the effect of SI by only 10%.

The corresponding reductions are 13% and 7% with additional regressors.
Within Estimators using Siblings or Twins

Within estimators based on siblings assume that omitted ability is family specific.

Let $i$ denote the family and $j$ denote the individual.

\begin{align*}
Y_{ij} &= \beta S_{ij} + \gamma A_i + u_{ij} \\
S_{ij} &= \eta A_i + v_{ij}
\end{align*}
Estimation:

One alternative is to difference the earnings of two siblings and thus eliminate unobserved family ability.

\[ dY = \beta dS + du \]

Assuming that the difference in schooling is orthogonal to the idiosyncratic differences in earnings, ols performed on the differenced equation will provide unbiased estimates of the schooling coefficient.
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma)A_i + u_{ij} + \beta v_{ij} \]

\[ S_{ij} = \eta A_i + v_{ij} \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma) A_i + u_{ij} + \beta v_{ij} \]

\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma^2_Y = (\beta \eta + \gamma)^2 \sigma^2_A + \beta^2 \sigma^2_v + \sigma^2_u \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma) A_i + u_{ij} + \beta v_{ij} \]

\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma^2_Y = (\beta \eta + \gamma)^2 \sigma^2_A + \beta^2 \sigma^2_v + \sigma^2_u \]

\[ \sigma^2_S = \eta^2 \sigma^2_A + \sigma^2_v \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma)A_i + u_{ij} + \beta v_{ij} \]
\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma_Y^2 = (\beta \eta + \gamma)^2 \sigma_A^2 + \beta^2 \sigma_v^2 + \sigma_u^2 \]
\[ \sigma_S^2 = \eta^2 \sigma_A^2 + \sigma_v^2 \quad \sigma_{YS} = \eta (\beta \eta + \gamma) \sigma_A^2 + \beta \sigma_v^2 \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma) A_i + u_{ij} + \beta v_{ij} \]

\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma_Y^2 = (\beta \eta + \gamma)^2 \sigma_A^2 + \beta^2 \sigma_v^2 + \sigma_u^2 \]

\[ \sigma_S^2 = \eta^2 \sigma_A^2 + \sigma_v^2 \]

\[ \sigma_{YS} = \eta (\beta \eta + \gamma) \sigma_A^2 + \beta \sigma_v^2 \]

\[ \sigma_{YY'} = (\beta \eta + \gamma)^2 \sigma_A^2 \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma) A_i + u_{ij} + \beta v_{ij} \]

\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma^2_Y = (\beta \eta + \gamma)^2 \sigma^2_A + \beta^2 \sigma^2_v + \sigma^2_u \]

\[ \sigma^2_S = \eta^2 \sigma^2_A + \sigma^2_v \]

\[ \sigma_{YS} = \eta (\beta \eta + \gamma) \sigma^2_A + \beta \sigma^2_v \]

\[ \sigma_{YY'} = (\beta \eta + \gamma)^2 \sigma^2_A \]

\[ \sigma_{SS'} = \eta^2 \sigma^2_A \]
A second alternative is to match moments. If we substitute the schooling equation, we get

\[ Y_{ij} = (\beta \eta_i + \gamma) A_i + u_{ij} + \beta v_{ij} \]
\[ S_{ij} = \eta A_i + v_{ij} \]

This yields the following second-order moments

\[ \sigma^2_Y = (\beta \eta + \gamma)^2 \sigma^2_A + \beta^2 \sigma^2_v + \sigma^2_u \]
\[ \sigma^2_S = \eta^2 \sigma^2_A + \sigma^2_v \]
\[ \sigma_{YS} = \eta (\beta \eta + \gamma) \sigma^2_A + \beta \sigma^2_v \]
\[ \sigma_{YY'} = (\beta \eta + \gamma)^2 \sigma^2_A \]
\[ \sigma_{SS'} = \eta \sigma^2_A \]
\[ \sigma_{YS'} = \eta (\beta \eta + \gamma) \sigma^2_A \]
Are all of the parameters identified?
Are all of the parameters identified?

There are 6 moments and 6 parameters:

$$ \beta, \eta, \gamma, \sigma^2_A, \sigma^2_v, \sigma^2_u $$

But, $\eta$ and $\sigma^2_A$ cannot be separately identified. One of them can be normalized to one.

The assumption that ability is a family specific endowment, the same for all children, is obviously very strong. Monozygotic twins are, however, identical in their genetic inheritance at conception.

The assumption that ability is a family specific endowment, the same for all children, is obviously very strong. Monozygotic twins are, however, identical in their genetic inheritance at conception.

Therefore, for them, the assumption that ability is family-specific is accurate to the extent that ability is captured by genetic endowment.
If twins are identical in genetic composition, then why did they obtain different levels of schooling?
If twins are identical in genetic composition, then why did they obtain different levels of schooling?

The fundamental identification assumption is that $\sigma_{uu} = 0$, that unobserved factors that affect the schooling decision are uncorrelated with unobserved factors that affect earnings.
If twins are identical in genetic composition, then why did they obtain different levels of schooling?

The fundamental identification assumption is that \( \sigma_{vu} = 0 \), that unobserved factors that affect the schooling decision are uncorrelated with unobserved factors that affect earnings.

There are a number of reasons to question that assumption.
1. Although monozygotic twins are genetically identical, they are not identical at birth. For example, twins differ in their birth weight, reflecting the fact that the environment within the womb is not identical. Innate ability, ability at birth, may thus be different from ability at conception.
1. Although monozygotic twins are genetically identical, they are not identical at birth. For example, twins differ in their birth weight, reflecting the fact that the environment within the womb is not identical. Innate ability, ability at birth, may thus be different from ability at conception.

2. Twins may experience different childhood events that affect their abilities later in life. For example, one of the twins may experience an accident that leads to a physical or mental impairment that affects schooling and earnings potential.
3. To the extent that schooling is affected by stochastic labor market opportunities, one twin may receive an offer at a high wage job that induces that twin to leave school.
3. To the extent that schooling is affected by stochastic labor market opportunities, one twin may receive an offer at a high wage job that induces that twin to leave school.

Although the first two of these possibilities lead to an upward biased estimate of the schooling coefficient, the last possibility leads to a downward biased estimate.
Instrumental Variables: Angrist and Krueger (1991)

The third approach to addressing the problem of ability bias in estimating schooling coefficients is to find an instrumental variable that is correlated with years of schooling but uncorrelated with ability.
Instrumental Variables: Angrist and Krueger (1991)

The third approach to addressing the problem of ability bias in estimating schooling coefficients is to find an instrumental variable that is correlated with years of schooling but uncorrelated with ability.

A well known study that applies this approach is that of Angrist and Krueger (1991), which uses state variation in school entry ages and in compulsory schooling laws as instruments for the amount of schooling completed.
Let $a_e$ be the minimum age at which a child is allowed to enter school.
Let $a_e$ be the minimum age at which a child is allowed to enter school.

Let $a_k$ denote the minimum age at which a child is allowed to drop out of school.
Let $a_e$ be the minimum age at which a child is allowed to enter school.

Let $a_k$ denote the minimum age at which a child is allowed to drop out of school.

Suppose that the last date that a child who turns age 6 can enter school is August 31 and that $a_k = 16$. 
Let $a_e$ be the minimum age at which a child is allowed to enter school.

Let $a_k$ denote the minimum age at which a child is allowed to drop out of school.

Suppose that the last date that a child who turns age 6 can enter school is August 31 and that $a_k = 16$.

Consider a child that is born on August 31. Then that child will enter school at exactly age 6 and will have completed 10 years of schooling upon reaching age 16.
Alternatively, consider a child that is born on September 1. That child will have to delay school entry until the following year, will enter school at age 7 and will complete only 9 years of schooling by age 16.
Alternatively, consider a child that is born on September 1. That child will have to delay school entry until the following year, will enter school at age 7 and will complete only 9 years of schooling by age 16.

If some people prefer to complete only 9 rather than 10 years of schooling, then average schooling will be lower for those born on September 1 relative to those born on August 31.
If birth date is uncorrelated with ability, then the difference in schooling will also be uncorrelated with ability.
If birth date is uncorrelated with ability, then the difference in schooling will also be uncorrelated with ability.

Let group 1 be those who did not meet the deadline (those born on September 1) and group 2 be those that did (those born on August 31).
If birth date is uncorrelated with ability, then the difference in schooling will also be uncorrelated with ability.

Let group 1 be those who did not meet the deadline (those born on September 1) and group 2 be those that did (those born on August 31).

Define a variable:
\[ z = 1 \text{ if an individual is in group 2} \]
\[ z = 0 \text{ if an individual is in group 1} \]
Consider the log wage function (where we ignore experience for now and where $Y$ represents log wage)

\[ Y_i = \beta S_i + \gamma A_i + u_i; \]
Consider the log wage function (where we ignore experience for now and where $Y$ represents log wage)

$$Y_i = \beta S_i + \gamma A_i + u_i;$$

The mean log wage for group 2 individuals is

$$E(Y_i|z_i = 1) = \beta E(S_i|z_i = 1) + \gamma E(A_i|z_i = 1) + E(u_i|z_i = 1)$$
Consider the log wage function (where we ignore experience for now and where $Y$ represents log wage)

$$Y_i = \beta S_i + \gamma A_i + u_i;$$

The mean log wage for group 2 individuals is

$$E(Y_i|z_i = 1) = \beta E(S_i|z_i = 1) + \gamma E(A_i|z_i = 1) + E(u_i|z_i = 1)$$

Similarly, for group 1 individuals, it is

$$E(Y_i|z_i = 0) = \beta E(S_i|z_i = 0) + \gamma E(A_i|z_i = 0) + E(u_i|z_i = 0).$$
Under the assumption that $z$ is uncorrelated with ability and the other unobserved determinants of schooling

\[ E(A_i | z_i = 1) = E(A_i | z_i = 0) \]

and

\[ E(u_i | z_i = 1) = E(u_i | z_i = 0). \]
Under the assumption that $z$ is uncorrelated with ability and the other unobserved determinants of schooling

\[ E(A_i | z_i = 1) = E(A_i | z_i = 0) \]

and

\[ E(u_i | z_i = 1) = E(u_i | z_i = 0). \]

Taking differences and solving for $\beta$, we get

\[ \beta = \frac{E(Y_i | z_i = 1) - E(Y_i | z_i = 0)}{E(S_i | z_i = 1) - E(S_i | z_i = 0)} \]
Replacing expectations with sample analogs gives the Wald estimate,

$$\hat{\beta} = \frac{\Delta Y}{\Delta S} = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{s}_2 - \bar{s}_1}.$$
Replacing expectations with sample analogs gives the Wald estimate,

$$\hat{\beta} = \frac{\Delta Y}{\Delta S} = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{s}_2 - \bar{s}_1}.$$  

Angrist and Krueger find that for men age 41-50 in 1970:

$$\bar{s}_2 - \bar{s}_1 = 0.1256 \quad \text{Wald estimate}=0.0715$$  

$$\text{ols estimate}=0.0801$$
Replacing expectations with sample analogs gives the Wald estimate,

\[ \hat{\beta} = \frac{\Delta Y}{\Delta S} = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{s}_2 - \bar{s}_1}. \]

Angrist and Krueger find that for men age 41-50 in 1970:

\[ \bar{s}_2 - \bar{s}_1 = 0.1256 \quad \text{Wald estimate} = 0.0715 \quad \text{ols estimate} = 0.0801 \]

They find for men in 1980:

\[ \bar{s}_2 - \bar{s}_1 = 0.1888 \quad \text{Wald estimate} = 0.102 \quad \text{ols estimate} = 0.0709 \]
What assumptions are required to interpret the Wald estimate as the schooling coefficient in the BP-G human capital pricing equation?

To see that, we need a model of the schooling decision.
Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.
Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.

2. Assume that there is only one decision period after reaching the compulsory schooling age, attend school or not attend school.
Assume:

1. that everyone works full-time for the same number of periods after leaving school so that actual work experience is the same as potential work experience.

2. Assume that there is only one decision period after reaching the compulsory schooling age, attend school or not attend school.

3. Assume that there is a direct cost of attending school in that decision period.
Define the attendance choice as

\[ s_1 \in \{0, 1\} \]
Define the attendance choice as

\[ s_1 \in \{0, 1\} \]

and let \( S_0 \) be the number of years of schooling completed at the compulsory school leaving age. Then, completed schooling is

\[ S_1 = S_0 + s_1 \]
Define the attendance choice as

\[ s_1 \in \{0, 1\} \]

and let \( S_0 \) be the number of years of schooling completed at the compulsory school leaving age. Then, completed schooling is

\[ S_1 = S_0 + s_1 \]

Let the wage function be separable in schooling and other determinants of skill (work experience, \( x \)), but not in ability, \( \mu \) :

\[ \log y = f(S, \mu) + g(x, \mu) \]
The present value of lifetime earnings for each schooling alternative is given by

\[
V(s_1 = 1|S_0) = \exp[f(S_0 + 1, \mu)] \sum_{x=0}^{X} \beta^{x+1} \exp[g(x, \mu)] - c
\]

\[
V(s_1 = 0|S_0) = \exp[f(S_0, \mu)] \sum_{x=0}^{X} \beta^{x} \exp[g(x, \mu)]
\]
The present value of lifetime earnings for each schooling alternative is given by

\[
V(s_1 = 1| S_0) = \exp[f(S_0 + 1, \mu)] \sum_{x=0}^{X} \beta^{x+1} \exp[g(x, \mu)] - c
\]

\[
V(s_1 = 0| S_0) = \exp[f(S_0, \mu)] \sum_{x=0}^{X} \beta^{x} \exp[g(x, \mu)]
\]

The school attendance decision is:

\[
s_1 = 1 \text{ if } f(S_0 + 1, \mu) - f(S_0, \mu) \geq r + \log \left[ \frac{c}{V(s_1 = 0| S_0)} + 1 \right]
\]

= 0 otherwise

where \( \beta = 1/(1 + r) \).
Note that:

1. If \( c = 0 \), then we obtain the usual condition that attendance depends on whether the marginal return exceeds the interest rate: \( \frac{\Delta \log y}{\Delta s} \geq r \).
Note that:

1. If \( c = 0 \), then we obtain the usual condition attendance depends on whether the marginal return exceeds the interest rate: \( \frac{\Delta \log y}{\Delta s} \geq r \).

2. If \( \frac{\partial}{\partial \mu} [f(S_0 + 1, \mu) - f(S_0, \mu)] > 0 \), then there exists a \( \mu^* \) such that \( s_1 = 1 \) if \( \mu \geq \mu^* \).
Given the condition in 2, those that attend school are of higher ability than those that do not attend.
Given the condition in 2, those that attend school are of higher ability than those that do not attend.

Thus, \( \frac{\Delta \log y}{\Delta s} \) as measured by the difference in earnings of the two schooling groups, will reflect ability differences.
Given the condition in 2, those that attend school are of higher ability than those that do not attend.

Thus, \( \frac{\Delta \log y}{\Delta s} \) as measured by the difference in earnings of the two schooling groups, will reflect ability differences.

The bias in the estimated schooling effect due to omitted (unobserved) ability is

\[
E \left( \frac{\Delta \log y}{\Delta s} \right) = E_\mu [f(S_0 + 1, \mu)|\mu \geq \mu^*] - E_\mu [f(S_0, \mu)|\mu < \mu^*]
\]

\[
> E_\mu [f(S_0 + 1, \mu) - f(S_0, \mu)]
\]
Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.
Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

$$\mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2.$$
Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

\[ \mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2. \]

The proportion of type 1 individuals is \( \pi_1 \) and the proportion of type 2’s \( 1 - \pi_1 \). Assume that the types are independent of date of birth (instrument validity).
Consider the interpretation of the Angrist and Krueger estimator within this schooling choice model.

Assume that there are only two ability types:

\[ \mu_1 \text{ and } \mu_2, \text{ with } \mu_1 > \mu_2. \]

The proportion of type 1 individuals is \( \pi_1 \) and the proportion of type 2’s \( 1 - \pi_1 \). Assume that the types are independent of date of birth (instrument validity).

Suppose the optimal level of schooling for type 1’s is \( S_0 + 1 \) and that of type 2’s \( S_0 \).
Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.
Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it’s optimal for them.
Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it’s optimal for them.

The older type 2 children complete $S_0$ years because that is optimal for them.
Compare two sets of children, those who just make the school entry date of birth deadline (older children) and those who just miss the deadline (younger children) – they differ in age, say, by only 1 day.

The type 1 children complete $S_0 + 1$ years regardless of their date of birth because it’s optimal for them.

The older type 2 children complete $S_0$ years because that is optimal for them.

But, the younger type 2 children are forced to remain in school an extra year because they reach the school leaving age only after spending $S_0 + 1$ years in school.
To get the Wald estimate of the schooling effect, note that:

Mean earnings for younger type 1’s \[ = f(S_0 + 1, \mu_1) \]

Mean earnings for younger type 2’s \[ = f(S_0 + 1, \mu_2) \]

Mean earnings for older type 1’s \[ = f(S_0 + 1, \mu_1) \]

Mean earnings for older type 2’s \[ = f(S_0, \mu_2) \]
Mean earnings of younger children

\[ = \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) \]

Mean earnings of older children

\[ = \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2) \]
Thus, the difference in mean earnings between the younger and older children is

\[ \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) - \pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2) \]
Thus, the difference in mean earnings between the younger and older children is

\[
\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) - \\
\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2) = (1 - \pi_1)[f(S_0 + 1, \mu_2) - f(S_0, \mu_2)]
\]
Thus, the difference in mean earnings between the younger and older children is

\[
\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0 + 1, \mu_2) - \\
\pi_1 f(S_0 + 1, \mu_1) + (1 - \pi_1) f(S_0, \mu_2)
\]

\[
= (1 - \pi_1)[f(S_0 + 1, \mu_2) - f(S_0, \mu_2)]
\]

and the change in the population mean schooling is

\[
\pi_1 0 + (1 - \pi_1) 1
\]
Thus, the Wald estimate is

\[
\frac{\Delta E(Y)}{\Delta E(S)} = f(S_0 + 1, \mu_2) - f(S_0, \mu_2).
\]
Thus, the Wald estimate is

$$\frac{\Delta E(Y)}{\Delta E(S)} = f(S_0 + 1, \mu_2) - f(S_0, \mu_2).$$

This is the marginal effect of schooling on earnings for the less able only.
When would the Wald estimate equal the marginal effect for the population (not just for the less able)?
When would the Wald estimate equal the marginal effect for the population (not just for the less able)?

\[ f(S, \mu) = f_1(S) + f_2(\mu) \]

i.e., if the marginal effect of schooling is independent of ability.
What happens when we allow individuals to make labor supply decisions after leaving school?
What happens when we allow individuals to make labor supply decisions after leaving school?

Let \( x_a \) denote work experience. Then, given

\[
\log y_a = f(s, \mu) + g(x_a, \mu),
\]

\[
\frac{\Delta \log y_a}{\Delta s} = \frac{\Delta f}{\Delta s} + \frac{\Delta g}{\Delta x} \frac{\Delta x}{\Delta s}.
\]
What happens when we allow individuals to make labor supply decisions after leaving school?

Let $x_a$ denote work experience. Then, given

$$\log y_a = f(s, \mu) + g(x_a, \mu),$$

$$\frac{\Delta \log y_a}{\Delta s} = \frac{\Delta f}{\Delta s} + \frac{\Delta g}{\Delta x} \frac{\Delta x}{\Delta s}.$$

The term $\frac{\Delta x}{\Delta s} > 0$ reflects the incentive of those with more schooling, who thus have higher future wages, to also work more in the future.