14.662, Spring 2015: Problem Set 4
Due Wednesday, May 6 (before class)
Heidi L. Williams
TA: Peter Hull (hull@mit.edu)

1. **Taste-Based Discrimination with Wage Posting**

   In recitation we saw how job search can magnify slight discriminatory preferences among a minority of firms. Here we show this in a model of targeted search over openings with pre-anounced (posted) wages, as considered by Lang et al. (2005).

   Suppose all potential workers in a given market are equally productive, generating value v to each of N identical firms. Each firm has one unfilled position and announces a wage to attract a single applicant. A random and unobservable (to both firms and workers) number $\tilde{Z}$ of workers observe all postings and each apply to only one job. It is common knowledge that $\tilde{Z}$ is distributed Poisson with mean Z (recall this implies that $P(\tilde{Z} = k) = Z^k \exp(-Z)/k!$). If more than one worker applies, the firm chooses randomly between them.

   Denote the posted wage of each firm i by w_i and the vector of all N wages by W. Workers play mixed strategies in choosing where to apply; write these by $q(W) = (q_1(W), ..., q_n(W))$ where $q_i(W)$ is the probability a worker will apply to firm i given announced wages. We assume anonymous strategies, so that for a W where $w_i = w_j$ we have $q_i(W) = q_j(W)$. We further restrict attention to symmetric equilibria in which all workers use the same strategy $q^*(\cdot)$ taking equilibrium wages as given. In such an equilibrium the number of workers applying to firm i will also be distributed Poisson with mean $z_i = q_i^*(W)Z$.

   (a) (6 points) *Write an expression for the probability that a worker will be hired by a firm facing an expected number of applicants z. Use this to write the expected payoff a worker expects from applying to firm i and characterize the firm's expected number of applicants. Show that market clearing then defines a unique symmetric equilibrium of the worker application subgame given W.* The probability a worker will be hired is

   $$f(z) = \sum_{n=0}^{\infty} \frac{1}{n+1} \frac{z^n \exp(-z)}{n!}$$

   where $1/(n+1)$ is the probability of hire given n other workers apply and $z^n \exp(-n)/n!$ is the Poisson probability of that case. The infinite sum may be shown to equal $f(z) = (1 - \exp(-z))/z$ for $z > 0$ (clearly $f(z) = 1$ if $z = 0$). The expected payoff from applying is then

   $$K_i = w_i f(z_i)$$

   In a symmetric equilibrium in which all workers play the same strategy, workers will only apply to firms with $K_i = \max_i\{K_i\} \equiv K$, which we can think of as the market expected income. In this equilibrium there will be no incentive for a firm to offer a wage higher than K, and if a firm offers a wage w_i less than or equal to K no worker will apply. That is,

   $$z_i > 0 \text{ for } w_i \geq K$$
   $$z_i = 0 \text{ for } w_i < K$$

Specifically, $z_i = f^{-1}(K/w_i)$ for $w_i \geq K$, and the total expected number of applicants may be written

$$\sum_{i=1}^{N} z_i = \sum_{\{i|w_i \geq K\}} f^{-1}(K/w_i)$$

In equilibrium this quantity must equal Z, the total expected supply of applicants. Since $f^{-1}(K/w_i)$ is strictly decreasing in K and the sum can only lose (positive) terms as K increases, this expression is strictly decreasing in K. Thus there exists a unique solution for K for a given wage schedule W. This solution may be used to yield the vector of equilibrium probabilities $q^*(W)$ from the fact that $q_i Z = z_i = f^{-1}(K/w_i)$.

(b) (5 points) *Write an expression for each firm's expected profits given w_i and z_i. Use this and your results in (a) to derive the optimal choice of z_i for a profit-maximizing firm that takes the structure of wages as given. What are equilibrium wages? What are equilibrium expected payoffs to workers and firms?* Firms expect profits of

$$\pi_i = (1 - \exp(-z_i))(v - w_i)$$

where $1 - \exp(-z_i)$ is the probability a firm fills its vacancy and $v - w_i$ is per-worker profit. Since in equilibrium an operating firm has $w_i = K/f(z_i)$ we can write

$$\begin{aligned} \pi_i &= (1 - \exp(-z_i))(v - K/f(z_i)) \\ &= (1 - \exp(-z_i))v - z_i K \end{aligned}$$

Profit-maximization then implies

$$\begin{aligned} &\exp(-z_i^*)v - K = 0 \\ &\implies z_i^* = \ln(\frac{v}{K}) \end{aligned}$$

Thus in equilibrium every firm expects the same number of applicants, $\ln(\nu/K)$. Since each worker applies to exactly one firm, market clearing implies $z_i^* = Z/N \equiv r$. Expected worker wages and income and firm profits may then be written

$$\begin{aligned} w^* &= \frac{vr}{\exp(r) - 1} \\ K^* &= v\exp(-r) \\ \pi^* &= (1 - (1 + r)\exp(-r))v \end{aligned}$$

(c) (5 points) *Now suppose there are two types of workers, black and white, the total numbers of which are distributed Poisson with means Z and Y respectively. The productivity of white workers remains ν, while the value of black workers to firms is $(1-\delta)\nu$, where the parameter δ (reflecting taste-based discrimination or actual physical differences in production) is small or zero. Assume even when $\delta = 0$ all firms find black workers to be marginally less desirable, so that they will always choose to hire a white worker when both types apply (but still choosing randomly within racial groups). As before each firm posts a single wage that they commit to paying regardless of race. Characterize the symmetric equilibrium strategy of white workers given wage postings W and knowledge of discriminatory hiring practices.*

White workers can effectively ignore the presence of black applicants, since they have no effect on the probability of their being hiring. Therefore the applications behavior of white workers is characterized exactly as above. Denoting the expected income of white workers by H the expected number of white applicants z_i to firm i is given by the solution of $w_i f(z) = H$ for firms posting wages $w_i \geq H$ and zero otherwise.

(d) (5 points) *Denote the expected number of white and black applicants to a firm i by z_i and y_i. What is the probability $g(y_i, z_i)$ that a black worker will be hired given z_i and y_i, and what is his expected income at such a firm? Use this to characterize a symmetric equilibrium strategy of black workers given wage postings and knowledge of discriminatory hiring practices.*

Black applicants will be hired only if no whites apply, a Poisson event that occurs with probability $\exp(-z_i)$. Given this the probability of being hired among the pool of black workers exactly parallels the above, so that

$$g(y, z) = \exp(-z) f(y)$$

and black workers' expected income at firm i is $J_i = w_i g(y_i, z_i)$. As before, we can argue that blacks will only apply to firms with positive probability if they can attain the maximum expected (black) income $J \equiv \max_i\{J_i\}$ and thus won't apply to any job posting with $w_i < J$. For sufficiently high wages, however, blacks also won't apply because the probability of whites also applying will be high enough to reduce their expected probability of hire to bring J_i below J. Denoting this wage threshold by $\hat{w}_i$, we have that $y_i = 0$ for $w_i < J$ or $w_i > \hat{w}_i$ and otherwise y_i solves $w_i g(y, z_i) = J$. From this we can see $J < H$, so that in expectation blacks earn less than whites.

(e) (6 points) *Write an expression for expected firm profits given w_i, z_i, and y_i. Using your expressions for the expected earnings of white and black workers, derive and sign an expression for $\partial z_i / \partial w_i + \partial y_i / \partial w_i$, the marginal change in the expected number of job applicants given an increase in wages. For arbitrarily small δ, argue that in equilibrium some firms will only attract whites ("white firms") while others will only attract blacks ("black firms"). Discuss.*

In general a firm's expected profits are

$$\pi(y_i, z_i, w_i) = (1 - \exp(-z_i))(\nu - w_i) + \exp(-z_i)(1 - \exp(-y_i))((1 - \delta)\nu - w_i)$$

where the first term is expected profits from white workers and the second is expected profits from black workers. Implicitly differentiating the expected earnings conditions we have

$$\begin{aligned} w_i f(z_i) &= H \\ \implies \frac{\partial z_i}{\partial w_i} &= -\frac{f(z_i)}{w_i f'(z_i)} > 0 \end{aligned}$$

and

$$\begin{aligned} w_i g(y_i, z_i) &= J \\ \implies \frac{\partial y_i}{\partial w_i} &= \frac{f(y)}{f'(y)} \left(\frac{\partial z_i}{\partial w_i} - \frac{1}{w_i} \right) \end{aligned}$$

so that

$$\frac{\partial z_i}{\partial w_i} + \frac{\partial y_i}{\partial w_i} = \frac{f(y)}{f'(y)} \left(\frac{\partial z_i}{\partial w_i} \right) < 0$$

By decreasing wages a firm decreases their probability of having a job vacancy while at the same time lowers their expected labor costs. For arbitrarily small δ this leads to an increase in expected profits. If a firm offers a wage that attracts both white and black workers, it would thus prefer to marginally decrease its wage and discourage more white applicants from applying. Therefore in equilibrium no firm would ever choose to attract both white and black workers, leading to complete segregation, even in the case where $\delta = 0$.

(f) (6 points) *Let N_z and N_y be the numbers of white and black firms with $r_z \equiv Z/N_z$ and $r_y \equiv Y/N_y$ denoting the expected number of applicants to each type of firm. Write expressions for equilibrium wages and expected profits of white firms and the expected income of white workers. Argue that equilibrium wages for black firms will be set at the expected income of white workers given arbitrarily small δ. Use this to derive expressions for the expected income of black workers and profits of black firms. Describe the discriminatory equilibrium. Are black workers unambiguously worse off than white workers? How does the expected income of white and black workers compare to the model without discrimination?*

White workers and firms will behave as in the nondiscriminatory model, so that

$$\begin{aligned} w_w^* &= \frac{\nu r_z}{\exp(r_z) - 1} \\ H^* &= v \exp(-r_z) \\ \pi_w^* &= (1 - (1 + r_z) \exp(-r_z))v \end{aligned}$$

Whites will not apply to any firm offering less than H^*. For sufficiently small δ, black wages should be such that a firm is just indifferent between posting w_b^* (becoming a black firm) and marginally increasing wages to attract a single white worker (becoming a white firm) by promising to pay her expected income of H^*. Thus $w_b^* = H^*$and since the expected number of black applicants must be equal to $r_y \equiv Y/N_y$ the equilibrium expected black income may be written

$$\begin{aligned} J^* &= w_b^* \exp(-z) f(y) \\ &= H^* \frac{1 - \exp(-r_y)}{r_y} \end{aligned}$$

Profits for black firms are then

$$\pi_b^* = (1 - \exp(-r_y))((1 - \delta)\nu - H^*)$$

Black workers are strictly worse off than white workers, since $f(r_y) < 1$. Furthermore both black and white workers face lower wages and expected incomes when firms are discriminatory, since $\rho/(\exp(\rho)-1)$ is decreasing in its argument and since $r_z = Z/N_z > Z/(N_z+N_y) = r$. Discrimination clearly hurts black workers, since they face a lower probability of being hired. However by lowering wages in the black sector, discrimination increases the profitability of hiring blacks when they are close substitutes in production. This induces more firms to set low wages that attract only black workers, reducing the demand for white workers and thus the wages that are posted by white firms. It can correspondingly be shown that firm profits are higher in the discriminatory regime. This dynamic is quite reminiscent of how firms behave monopsonistically in the search model of Black (1995), though the ultimate source of this market power is quite different.

2. **Wage Discrimination with Endogenous Human Capital Investment**

Many of the models we've seen take groups' human capital investments as given and analyze the implications for wage disparities. Lundberg and Startz (1983) develop a model in which workers anticipate wage setting practices in making their human capital investments. Here you'll analyze the implications of such "endogenous discrimination" for the observed skill/wage distribution.

Let π_i denote the productivity of worker i, which depends on innate ability a_i and acquired skill X_i.

$$\pi_i = a_i + bX_i$$

Such skill is acquired at cost

$$c(X_i) = \left(\frac{c}{2}\right) X_i^2$$

for $c > 0$. Both workers and employers know the values of b and c, but employers do not observe true productivity. Instead, they observe test scores that measure productivity with independent error ϵ_i.

$$T_i = \pi_i + \varepsilon_i$$

Both workers and employers know a_i and ε_i are normally distributed with means $\bar{a}$ and $\bar{\varepsilon}$ and with variances σ_a^2, and σ_ε^2.

(a) (5 points) *How much training would workers purchase if employers could observe productivity directly? What would be the equilibrium expense on training in this case?*

When π_i is perfectly observed, workers are paid their marginal product and solve

$$\max_x \quad a_i + bx - \left(\frac{c}{2}\right) x^2$$

The first-order condition implies that each worker purchases b/c units of training, paying $b^2/2c$.

(b) (6 points) *Solve for the equilibrium wage schedule under imperfect information. Start by assuming the optimal level of training for worker i may be written as a linear function of a_i and ε_i:*

$$X_i = \rho_0 + \rho_a a_i + \rho_\varepsilon \varepsilon_i$$

i. *Use this expression to derive employers' wage offers w_i in terms of observed test scores T_i the mean test score $\bar{T}$, and mean productivity $\bar{\pi}$.*

ii. *Taking this wage schedule as given, solve for the level of training that worker i acquires and interpret your result.*

iii. *Use your results from (i) and (ii) to derive a new expression for the wage schedule as a function of individual worker characteristics a_i and ε_i. Interpret your result.*

With imperfect information, wages equal expected productivity given the observed test score. Because π_i and T_i are jointly normal:

$$\begin{aligned} w_i &= E[\pi_i \mid T_i] \\ &= \bar{\pi} + \frac{Cov(\pi_i, T_i)}{var(T_i)}(T_i - \bar{T}) \\ &= \bar{\pi} + \frac{Cov(a_i + bX_i, a_i + bX_i + \varepsilon_i)}{Var(a_i + bX_i + \varepsilon_i)}(T_i - \bar{T}) \end{aligned}$$

Substituting in for the linear expression for X_i, we have

$$\begin{aligned} Cov(a_i + bX_i, a_i + bX_i + \varepsilon_i) &= Var(a_i) + 2bCov(X_i, a_i) + b^2Var(X_i) + bCov(X_i, \varepsilon_i) \\ &= \sigma_a^2 + 2b\rho_a\sigma_a^2 + b^2(\rho_a^2\sigma_a^2 + \rho_\varepsilon^2\sigma_\varepsilon^2) + b\rho_\varepsilon\sigma_\varepsilon^2 \\ &= (1 + b\rho_a)^2\sigma_a^2 + b\rho_\varepsilon(1 + b\rho_\varepsilon)\sigma_\varepsilon^2 \end{aligned}$$

and

$$Var(a_i + bX_i + \varepsilon_i) = (1 + b\rho_a)^2\sigma_a^2 + (1 + b\rho_\varepsilon)^2\sigma_\varepsilon^2$$

Defining

$$B \equiv \frac{(1 + b\rho_a)^2\sigma_a^2 + b\rho_\varepsilon(1 + b\rho_\varepsilon)\sigma_\varepsilon^2}{(1 + b\rho_a)^2\sigma_a^2 + (1 + b\rho_\varepsilon)^2\sigma_\varepsilon^2}$$

We have workers solving

$$\max_x \quad \bar{\pi} + B(a_i + bx + \varepsilon_i - \bar{T}) - \left(\frac{c}{2}\right)x^2$$

and choosing $X_i = B(b/c)$. Since this expression does not vary with a_i or ε_i, we can conclude that $\rho_a = \rho_\varepsilon = 0$. Thus:

$$B = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2}$$

so that

$$X_i = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2}\left(\frac{b}{c}\right)$$

The wage schedule is then

$$\begin{aligned} w_i &= \bar{\pi} + \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2}\left(T_i - \bar{T}\right) \\ &= \bar{\pi} + \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\varepsilon^2}(a_i - \bar{a} + \varepsilon_i - \bar{\varepsilon}) \end{aligned}$$

Under perfect information, workers are rewarded 1:1 for every unit increase in their true productivity. When firms can't observe true productivity, they can't be sure whether above average test scores reflects higher productivity $(a_i - \bar{a})$ or simply measurement error $(\varepsilon_i - \bar{\varepsilon})$, so they shrink their observation back toward the mean using the familiar signal-to-noise ratio, B. Note that this lowers the equilibrium return to training, so that $X_i^{(a)} > X_i^{(b)}$. Consequently, imperfect information reduces the aggregate output of the economy.

(c) (5 points) *Now suppose that workers belong to two observable groups. The groups have identical mean innate characteristics $\bar{a}$ and $\bar{\varepsilon}$ and test variance σ_T^2, but group 1 has relatively heterogeneous innate ability and relatively homogenous testing ability. Formally, $\sigma_{a,1}^2 > \sigma_{a,2}^2$ and $\sigma_{\varepsilon,1}^2 < \sigma_{\varepsilon,2}^2$ for groups 1 and 2. Repeat part (b), this time allowing the wage schedules to differ by group. How much training does a worker in each group receive and what is her wage?*

Following the above derivation, the equilibrium training level for each group g is

$$X_g = \frac{\sigma_{a,g}^2}{\sigma_{a,g}^2 + \sigma_{\varepsilon,g}^2}\left(\frac{b}{c}\right) = \frac{\sigma_{a,g}^2}{\sigma_T^2}\left(\frac{b}{c}\right)$$

and the wage schedule is

$$w_i = \bar{\pi} + \frac{\sigma_{a,g(i)}^2}{\sigma_T^2}\left[(a_i - \bar{a}) + (\varepsilon_i - \bar{\varepsilon})\right]$$

Since the two groups share the same variance in test scores, all of the variation in wages and training comes from variation in innate ability. The group with more heterogeneous innate ability enjoys a higher return to test scores and, by extension, training, so they will invest more in human capital. The intuition is that when the total variance of test scores is the same across groups, test scores are more informative for the group for which variation in innate ability accounts for a larger share of the total score variance.

(d) (5 points) *Compare average wages for the two groups. Is this a discriminatory equilibrium by the Aigner and Cain standard? Why do Lundberg and Startz consider it discriminatory?*

The average wage for group g equals its average productivity:

$$\bar{w}_g = \bar{a} + \frac{\sigma_{a,g}^2}{\sigma_T^2}\left(\frac{b^2}{c}\right)$$

Since employers offer equal pay for equal expected productivity, this is not a discriminatory equilibrium by the Aigner and Cain standard. However, we find different wages for groups with the same innate ability, which is why Lundberg and Startz consider this a discriminatory outcome. Discrimination in this model stems from differences in groups' measured productivity, not differential endowments or discriminatory preferences among employers.

(e) (6 points) *Now suppose that employers are prohibited from offering group-specific wage schedules. Let $f(\varepsilon_i, T_i)$ denote the joint density of test-specific ability and test scores, and let f_1 and f_2 denote the corresponding group-specific densities. Let α denote the population fraction of group 1 workers so that*

$$f(\varepsilon_i, T_i) \quad = \quad \alpha f_1(\varepsilon_i, T_i) + (1-\alpha) f_2(\varepsilon_i, T_i)$$

i. *Assume that the equilibrium wage schedule will be linear in T_i, as in part (b), and let β denote the coefficient on T_i so that $w_i = \gamma + \beta T_i$ for some γ and β. Derive the optimal level of human capital investment as a function of β. Use your result to argue that f_1 and f_2 are bivariate normal.*

ii. *Derive an expression for* $f(\varepsilon_i \mid T_i)$ *and use it to derive the new wage schedule.*

iii. *How do average wages for the two groups compare now? Is this equilibrium discriminatory by the Aigner and Cain standard? By Lundberg and Startz's definition?*

Given a wage schedule that's linear in T_i, the worker solves

$$\max_x \quad \gamma + \beta\left(a_i + bx + \varepsilon_i\right) - \left(\frac{c}{2}\right)x^2$$

which again implies optimal training of $X_i = \beta\left(\frac{b}{c}\right)$. Note that X_i does not vary with a_i or ε_i, so test scores for group g are distributed

$$T_{i,g} \sim N(\bar{a} + \beta\left(\frac{b^2}{c}\right) + \bar{\varepsilon}, \sigma_T^2)$$

and $T_{i,g}$ and $\varepsilon_{i,g}$ are jointly normal with covariance σ_ε^2.
Note that we can write the conditional expectation of ϵ_i given test scores as

$$\begin{aligned}
E[\varepsilon_i \mid T_i] &= \alpha E[\varepsilon_i \mid T_i, g=1] + (1-\alpha)E[\varepsilon_i \mid T_i, g=2] \\
&= \alpha(\bar{\epsilon} + \frac{\sigma_{\varepsilon,1}^2}{\sigma_T^2}(T_i - \bar{T})) + (1-\alpha)\left(\bar{\epsilon} + \frac{\sigma_{\varepsilon,2}^2}{\sigma_T^2}(T_i - \bar{T})\right) \\
&= \bar{\varepsilon} + \left(\frac{\alpha\sigma_{\varepsilon,1}^2 + (1-\alpha)\sigma_{\varepsilon,2}^2}{\sigma_T^2}\right)(T_i - \bar{T}) \\
&= \bar{\varepsilon} + \left(1 - \frac{\alpha\sigma_{a,1}^2 + (1-\alpha)\sigma_{a,2}^2}{\sigma_T^2}\right)(T_i - \bar{T}) \\
&= \bar{\varepsilon} + (1-\bar{\beta})(T_i - \bar{T})
\end{aligned}$$

where $\bar{\beta} \equiv \alpha\left(\frac{\sigma_{a,1}^2}{\sigma_T^2}\right) + (1-\alpha)\left(\frac{\sigma_{a,2}^2}{\sigma_T^2}\right)$. It follows that the equilibrium wage schedule is

$$\begin{aligned}
w_i &= E[\pi_i|T_i] \\
&= E[T_i - \varepsilon_i|T_i] \\
&= T_i - \bar{\varepsilon} - (1-\bar{\beta})(T_i - \bar{T}) \\
&= \underbrace{-(\bar{\varepsilon} + \bar{\beta}\bar{T})}_{\gamma} + \bar{\beta}T_i
\end{aligned}$$

Note that average wages for the two groups are the same under this schedule because they share the same mean values of $\bar{T}$ and $\bar{\varepsilon}$. This outcome is not discriminatory by the Lundberg and Startz standard because workers with the same endowments receive the same wages, nor is it discriminatory by the Aigner and Cain standard because groups with the same average productivity receive the same pay on average. Note that an alternative interpretation of the "Aigner and Cain standard" is that workers with the same expected productivity given available signals are paid the same average wages, in which case this outcome is discriminatory.

(f) (6 points) *Compare the total amount of training obtained under each equilibrium.*

- *Perfect information*
- *Imperfect information with group-specific wages*

- *Imperfect information with a common wage schedule*

How does the ban on group-specific wages affect social welfare? Why?

Assuming a mass one of individuals, total training obtained under each equilibrium is

equilibrium	total training
Perfect information	$\frac{b}{c}$
Imperfect info w/ group-specific wages	$\bar{\beta}\left(\frac{b}{c}\right)$
Imperfect info w/ common wage schedule	$\bar{\beta}\left(\frac{b}{c}\right)$

Total training is the same under each of the imperfect information equilibria and below the full information level. Notice, however, that social welfare is higher when group-specific wages are banned because the total cost of training is lower.

$$c\left(X^{(c)}\right) = \left(\frac{c}{2}\right)\left(\alpha\left(\left(\frac{\sigma_{a,1}^2}{\sigma_T^2}\right)\frac{b}{c}\right)^2 - (1-\alpha)\left(\left(\frac{\sigma_{a,2}^2}{\sigma_T^2}\right)\frac{b}{c}\right)^2\right)$$

$$c\left(X^{(e)}\right) = \left(\frac{c}{2}\right)\left(\alpha\left(\frac{\sigma_{a,1}^2}{\sigma_T^2}\right)\frac{b}{c} + (1-\alpha)\left(\frac{\sigma_{a,2}^2}{\sigma_T^2}\right)\frac{b}{c}\right)^2$$

Since $f(x) = x^2$ is a convex function, we know $c\left(X^{(e)}\right) < c\left(X^{(c)}\right)$. The intuition is that the group-specific wage ban generates the same output but induces more training from the group that can acquire it at a lower cost.

3. **Intergenerational Mobility**

Consider a modified formalization of the Becker and Tomes (1979) model that we covered in class. Family i consists of a parent of generation $t-1$ that allocates earnings $y_{i,t-1}$ between consumption $C_{i,t-1}$ and investment $I_{i,t-1}$ in the human capital of her child of generation t. The investment technology is given by

$$h_{it} = \delta + \theta \ln I_{i,t-1} + e_{it}$$

for $\theta > 0$ and other sources of human capital (including genetics or cultural inheritance) e_{it}. As with Becker and Tomes (1979) suppose e_{it} is first-order autoregressive:

$$e_{it} = \lambda e_{i,t-1} + \nu_{it}$$

for $\lambda \in (0,1)$ and a white noise error term ν_{it}. The child's lifetime income is given by

$$\ln y_{it} = \mu + \rho h_{it}$$

where $\rho > 0$ denotes the returns to human capital investment.

(a) (8 points) *Suppose parental utility is Cobb-Douglas in consumption and childhood earnings:*

$$U_{i,t-1} = C_{i,t-1}^{1-\alpha} y_{it}^{\alpha}$$

Derive the optimal level of human capital investment and interpret. Taking logs of $U_{i,t-1}$ and substituting in, we have

$$\begin{aligned}
\ln U_{i,t-1} &= (1-\alpha)\ln C_{i,t-1} + \alpha \ln y_{it} \\
&= (1-\alpha)\ln(y_{i,t-1} - I_{i,t-1}) + \alpha(\mu + \rho h_{it}) \\
&= (1-\alpha)\ln(y_{i,t-1} - I_{i,t-1}) + \alpha\mu + \alpha\rho\delta + \alpha\rho\theta \ln I_{i,t-1} + \alpha\rho e_{it}
\end{aligned}$$

The first-order condition characterizing optimal investment is

$$-\frac{1-\alpha}{y_{i,t-1} - I_{i,t-1}} + \frac{\alpha\rho\theta}{I_{i,t-1}} = 0$$

$$\implies I_{i,t-1} = \frac{\alpha\rho\theta}{(1-\alpha) + \alpha\rho\theta} y_{i,t-1}$$

From this expression we can see that parents spend a constant fraction of their income on their child's education; this fraction is increasing in the level of "altruism," α, as well as $\rho\theta$, the earnings return to human capital investment; parents are more willing to invest in their children when the payoff to that investment is higher.

(b) (9 points) *Write an expression linking $\ln y_{it}$ and $\ln y_{i,t-1}$. Can a standard intergenerational income regression recover this relationship? Derive an expression (in terms of ρ, θ, and λ) for the intergenerational income elasticity coefficient produced under this model. Discuss.*

Substituting, we have

$$\begin{aligned}
\ln y_{it} &= \mu + \rho\delta + \rho\theta \ln I_{i,t-1} + \rho e_{it} \\
&= \mu + \rho\delta + \rho\theta \ln\left(\frac{\alpha\rho\theta}{(1-\alpha) + \alpha\rho\theta}\right) + \rho\theta \ln y_{i,t-1} + \rho e_{it}
\end{aligned}$$

Given the current model, a regression of $\ln y_{it}$ on $\ln y_{i,t-1}$ will not produce a coefficient of $\rho\theta$, because both the child's endowment e_{it} and the parent's log income $\ln y_{i,t-1}$ depends on the parent's endowment $e_{i,t-1}$. By repeated substitution we can write the regression coefficient as

$$\begin{aligned}
\beta &= \frac{Cov(\ln y_{it}, \ln y_{i,t-1})}{Var(\ln y_{i,t-1})} \\
&= \frac{Cov(\rho\theta \ln y_{i,t-1} + \rho(\lambda e_{i,t-1} + \nu_{it}), \ln y_{i,t-1})}{Var(\ln y_{i,t-1})} \\
&= \rho\theta + \frac{\rho\lambda Cov(e_{i,t-1}, \ln y_{i,t-1})}{\sigma_y^2} \\
&= \rho\theta + \frac{\rho\lambda Cov(e_{i,t-1}, \rho\theta \ln y_{i,t-2} + \rho e_{i,t-1})}{\sigma_y^2} \\
&= \rho\theta + \frac{\rho^2\lambda\sigma_e^2 + \rho\lambda\theta\left(\rho\lambda Cov(e_{i,t-2}, \ln y_{i,t-2})\right)}{\sigma_y^2} \\
&\vdots \\
&= \rho\theta + \frac{\rho^2\lambda\sigma_e^2}{(1-\rho\lambda\theta)\sigma_y^2}
\end{aligned}$$

Furthermore, we have that

$$\begin{aligned}
Cov(\ln y_{it}, \ln y_{it}) &= \rho\theta Cov(\ln y_{it}, \ln y_{i,t-1}) + \rho Cov(\ln y_{it}, e_{it}) \\
\sigma_y^2 &= \rho^2\theta^2\sigma_y^2 + \rho\theta\frac{\rho^2\lambda\sigma_e^2}{1-\rho\lambda\theta} + \frac{1}{\lambda}\frac{\rho^2\lambda\sigma_e^2}{1-\rho\lambda\theta} \\
\sigma_y^2 &= \sigma_e^2\frac{1}{1-\rho^2\theta^2}\left(\frac{\rho^3\lambda\theta}{1-\rho\lambda\theta} + \frac{\rho^2\lambda}{\lambda-\rho\lambda^2\theta}\right) \\
&= \sigma_e^2\frac{1}{1-\rho^2\theta^2}\left(\frac{\rho^2-\rho^4\lambda^2\theta^2}{(1-\rho\lambda\theta)^2}\right) \\
&= \sigma_e^2\frac{\rho^2}{1-\rho^2\theta^2}\left(\frac{1+\rho\lambda\theta}{1-\rho\lambda\theta}\right)
\end{aligned}$$

Thus we can write the estimated intergenerational elasticity as

$$\begin{aligned}
\beta &= \rho\theta + \frac{\rho^2\lambda\sigma_e^2}{(1-\rho\lambda\theta)\sigma_e^2\frac{\rho^2}{1-\rho^2\theta^2}\left(\frac{1+\rho\lambda\theta}{1-\rho\lambda\theta}\right)} \\
&= \rho\theta + \frac{\lambda-\rho^2\lambda\theta^2}{1+\rho\lambda\theta} \\
&= \frac{\rho\theta+\lambda}{1+\rho\lambda\theta}
\end{aligned}$$

From this expression, we can see that we'd get a positive regression coefficient both because the earnings return to human capital investment $\rho\theta$ is positive *and* because λ is positive (that is, parents with more favorable non-investment endowments tend to pass these on to their children). β will always overstate the true return $\rho\theta$ provided $\rho\theta < 1$ since

$$\frac{\rho\theta+\lambda}{1+\rho\lambda\theta} - \rho\theta \propto \lambda(1-\rho^2\theta^2)$$

(c) (8 points) *Suppose we have data on three generations of individuals: t (children), $t-1$ (parents), and $t-2$ (grandparents). Write an expression relating $\ln y_{it}$ to $\ln y_{i,t-1}$ and $\ln y_{i,t-2}$. Becker and Tomes (1979) were the first to note that in a regression of these variables the coefficient on log grandparent income can be negative (though small). A naïve observer of this fact might take this as refutation of a model where parents invest altruistically in their children. What is the correct interpretation of this finding?* Writing $c \equiv \mu + \rho\delta + \rho\theta \ln\left(\frac{\alpha\rho\theta}{(1-\alpha)+\alpha\rho\theta}\right)$,

we have

$$\begin{aligned} \ln y_{i,t-1} &= c + \rho\theta \ln y_{i,t-2} + \rho e_{i,t-1} \\ \ln y_{it} - \lambda \ln y_{i,t-1} &= (1-\lambda)c + \rho\theta \ln y_{i,t-1} - \lambda\rho\theta \ln y_{i,t-2} + \rho e_{i,t} + \lambda\rho e_{i,t-1} \\ \ln y_{it} &= (1-\lambda)c + (\rho\theta + \lambda)\ln y_{i,t-1} - \lambda\rho\theta \ln y_{i,t-2} + \rho\nu_{it} \end{aligned}$$

This expression characterizes the regression of $\ln y_{it}$ on $\ln y_{i,t-1}$ and $\ln y_{i,t-2}$, since $\rho\nu_{it}$ is a white-noise residual. From this we can see the small negative coefficient on $\ln y_{i,t-2}$ estimated by Becker and Tomes (1979) is exactly what is predicted by the model. The naïve observer is (naïvely) taking this to mean that an exogenous increase in grandparental income harms a child's income, but this is not the correct interpretation as such an increase would also cause parental income to rise. If the parent did not earn more despite the advantages of increased grandparental income, it would have to be because they drew poorly on their endowment term, which is partly passed on to the child.

(d) (8 points) *Now suppose endowments are more persistent across generations; that is, suppose e_{it} evolves as*

$$e_{it} = \lambda_1 e_{i,t-1} + \lambda_2 e_{i,t-2} + \nu_{it}$$

for $0 \le \lambda_2 < \lambda_1 < 1$. Write an equation linking a child's log income to his parent's, grandparent's, and great-grandparent's. How might the availability of great-grandparent income data update your priors on the parameters? Repeating similar steps to part (c), we now

have

$$\ln y_{it} = (1-\lambda_1-\lambda_2)c + (\rho\theta + \lambda_1)\ln y_{i,t-1} + (\lambda_2 - \rho\theta\lambda_1)\ln y_{i,t-2} - \lambda_2\rho\theta \ln y_{i,t-3} + \rho\nu_{it}$$

Now the regression coefficient on great-grandparental income is negative, for similar logic as in (c). The coefficient on grandparental income may be positive if $\lambda_2 > \lambda_1\rho\theta$, or when grandparental contributions to genetic or cultural inheritance are large enough to dominate the negative effect discussed in (c). If we find a negative coefficient on grandparental income when we run this regression, we won't be able to tell whether the endowment error structure is AR-1 or AR-2, but testing whether the coefficient on great-grandparental income is nonzero will rule out the AR-1 case. Note that this test is likely to be very weak, as $\lambda_2\rho\theta$ may be quite small.

4. **Early Life Determinants of Long-Run Outcomes**

Deming (2009) is interested (in part) in estimating the effect of Head Start participation on later student test scores. Participation in Head Start, however, is non-random. Altonji, Elder and Taber (2005) provide a framework for thinking about the magnitude of bias induced by selection.

(a) (8 points) *Using the dataset provided, estimate the effect of Head Start participation on standardized test scores, conditional on year fixed effects and the covariates provided in the dataset. Report your estimate of the effect of Head Start. Discuss some reasons you might expect this estimate to be biased.*

Deming (2009) Head Start replication

	(1)	(2)	(3)	(4)
Head Start	-0.104**	-0.050	-0.007	0.031
	(0.053)	(0.053)	(0.052)	(0.051)
Other preschool	0.242***	0.211***	0.149***	0.113**
	(0.055)	(0.054)	(0.054)	(0.052)
K	0.010	0.013	0.003	-0.017
Baseline controls	N	Y	Y	Y
Income controls	N	N	Y	Y
All other pre-treatment controls	N	N	N	Y

Notes: baseline controls include fixed effects for race, sex, year, and test age. Income controls include log income interacted by a dummy for age=3. The sample size is 1,967. Robust standard errors are reported in parentheses.

The above table presents OLS estimates of the effect of Head Start relative to no preschool on cognitive test scores. Column (1) has no additional controls and produces a negative and statistically significant estimate. This negative effect shrinks as we control for additional pre-treatment characteristics and turns positive in the full covariate specification, but the point estimate is not statistically significant. These results suggest that Head Start enrollees are negatively selected on test scores but that they do not perform worse than observably similar non-preschool students. We may still be concerned that the included covariates do not adequately control for unobserved determinants of test scores, such as parental investments in informal education, in which case these estimates would be biased.

(b) (9 points) *Now we will think about the process of selection into Head Start. Suppose that we can write the following structural equation:*

$$\begin{aligned} Y^* &= \alpha HS + W'\Gamma \\ &= \alpha HS + X'\gamma + \epsilon \end{aligned}$$

Here, α is the true causal effect of Head Start on test scores Y^*. W is the full set of variables (observed and unobserved) that determine Y^* along with HS, and Γ is the causal effect of W on Y^*. X is a vector of the observable components of W, and γ and ϵ are defined *so that $Cov(X, \epsilon) = 0$. These same X's may influence whether a student participates in Head Start. Write:*

$$HS = X'\beta + \tilde{HS}$$

where $X'\beta$ is the predicted value of HS from a regression of HS on X and $\tilde{HS}$ is the residual. Show that the coefficient on HS from a regression of Y^* on HS and X can be expressed as:

$$\tilde{\alpha} = \alpha + \frac{Var(HS)}{Var(\tilde{HS})}(E(\epsilon|HS=1) - E(\epsilon|HS=0))$$

Which parts of this expression are observable? Which are not?

The estimated coefficient on HS is

$$\begin{aligned}
\tilde{\alpha} &= \frac{Cov(Y^*, \tilde{HS})}{Var(\tilde{HS})} \\
&= \frac{Cov(\alpha[X'\beta + \tilde{HS}] + X'\gamma + \epsilon, \tilde{HS})}{Var(\tilde{HS})} \\
&= \alpha + \frac{Cov(\epsilon, \tilde{HS})}{Var(\tilde{HS})} \\
&= \alpha + \frac{E[\epsilon \cdot (HS - E[HS])]}{Var(\tilde{HS})} \\
&= \alpha + \frac{E[\epsilon \cdot (HS - E[HS])|HS=1]E[HS] + E[\epsilon \cdot (HS - E[HS])|HS=0](1 - E[HS])}{Var(\tilde{HS})} \\
&= \alpha + \frac{E[\epsilon|HS=1]E[HS](1 - E[HS]) - E[\epsilon \cdot |HS=0]E[HS](1 - E[HS])}{Var(\tilde{HS})} \\
&= \alpha + \frac{Var(HS)}{Var(\tilde{HS})}(E(\epsilon|HS=1) - E(\epsilon|HS=0))
\end{aligned}$$

We can compute $Var(HS)$ and $Var(\tilde{HS})$, but the nature of selection $(E(\epsilon|HS=1) - E(\epsilon|HS=0))$ is unobserved.

(c) (8 points) *Assume the following variant of Condition 4 in Altonji, Elder and Taber (2005):*

$$\frac{E[\epsilon|HS=1] - E[\epsilon|HS=0]}{Var(\epsilon)} = K\frac{E[X'\gamma|HS=1] - E[X'\gamma|HS=0]}{Var(X'\gamma)}$$

Describe intuitively what the term K is. Under the null that Head Start has no impact (i.e. $\alpha = 0$), derive an expression for $\tilde{\alpha}$ that depends only on K and observables. Assume (without loss) that $Var(\epsilon) = 1$.

K indicates the extent to which selection on observables predicts selection on unobservables. Under the null that $\alpha = 0$ and with $\text{Var}(\epsilon) = 1$, we have

$$\begin{aligned}
\tilde{\alpha} &= \frac{Var(HS)}{Var(\tilde{HS})}(E(\epsilon|HS=1) - E(\epsilon|HS=0)) \\
&= K\left(\frac{Var(HS)}{Var(\tilde{HS})Var(X'\gamma)}\right)(E[X'\gamma|HS=1] - E[X'\gamma|HS=0])
\end{aligned}$$

(d) (8 points) *Use your expression in part (c) to estimate K. How large would selection on unobservables relative to selection on observables have to be in order for the estimated effect of Head Start to come entirely from selection? Be sure to think carefully about how the covariates you include in X affect the sign of your estimates.*

The table in part (a) also presents estimates for K, using the above formula. The full covariate specification generates a small, negative estimate of -0.017. The sign suggests that children who enroll in Head Start are negatively selected on unobserved determinants of test scores conditional on all of the other covariates that we include (note that this is also consistent with the direction our estimated coefficient changes with more controls). That implies that our estimated effect of Head Start in Column (3) understates the true effect. The small magnitude suggests that selection on unobservables has a small effect on the estimate in Column (1) relative to selection on observables.

References

Altonji, Joseph, Todd Elder, and Christopher Taber, "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 2005, *113* (1), 151–184. 4, 4c

Becker, Gary and Nigel Tomes, "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy*, 1979, *87* (6), 1153–1189. 3, 3c

Deming, David, "Early childhood intervention and life-cycle skill development," *American Economic Journal: Applied Economics*, 2009, *1* (3), 111–134. 4

Lang, Kevin, Michael Manove, and William Dickens, "Racial discriminaiton in labor markets with posted wage offers," *American Economic Review*, 2005, *95* (4), 1327–1340. 1

Lundberg, Shelly and Richard Startz, "Private discrimination and social intervention in competitive labor markets," *American Economic Review*, 1983, *73* (3), 340–3347. 2

14.662 Labor Economics II
Spring 2015