Online Appendix to "Social Effects in Employer Learning: An Analysis of Siblings"

Neel Rao
University at Buffalo, SUNY

November 3, 2015

Abstract

The online appendix to the paper is organized as follows. Appendix A discusses the procedure for estimating the employer learning models in the main text. Appendix B presents a simple model of employee referrals. Appendix C derives welfare and policy implications. Additional empirical analyses referenced in the main text are included at the end of the document. The tables display results related to job search patterns, joint work-wage outcomes, inexperienced siblings, geographic mobility, economic distance, human capital measures, and non-wage outcomes.

A Empirical Implementation of Model

This appendix resolves some issues concerning the estimation of the employer learning models in the main text. A potential obstacle to implementing the tests in the paper is that the regression coefficients are predicted to change with age in the conditional expectation of one's log wage given one's own and a sibling's test scores and schooling. The analysis treated the ages of the siblings in each family as being fixed. In the data, however, siblings from a sample of households are interviewed over multiple years, and the age structure varies across families and over time. One way to deal with this problem might simply be to include interactions of schooling and test scores with age when estimating the conditional expectation function. Nonetheless, this approach is unattractive in the current setting, because the social learning model implies that the coefficients on test scores are a function not only of one own's age but also of a sibling's age. Hence, the number of interaction terms that would need to be included in the specification is an order of magnitude greater than that required under individual learning, making it difficult to obtain precise estimates for the coefficients of interest.

Remarkably, there is a simple procedure that in large part overcomes this estimation problem. First, I show that the main predictions of both the individual and the social learning model hold in aggregate. Specifically, if one considers all the pairs of younger and older siblings in a sample of sibships with different age structures, then the predictions of the two employer learning models for the coefficients on test scores also apply to the expected values of these coefficients for a randomly selected pair of siblings. This finding is somewhat surprising because these predictions involve a nonlinear function of the regression coefficients: the ratio of the coefficient on a sibling's test score to that on one's own test score. Nevertheless, the normality assumptions in this paper impose sufficient structure on the learning process to make aggregation of this sort possible. Second, I show that the pooled ordinary least squares estimator of the conditional expectation function will under reasonable conditions generate a consistent estimate of the expected values of the regression coefficients for a randomly selected pair of siblings, provided that one controls sufficiently flexibly for the ages of the siblings.

The details of the estimation procedure are as follows. To simplify the exposition, I assume that all families consist of exactly two siblings and that all sibships enter the labor market in the same year. Consider a random sample of $I \geq 1$ sibships. The families in the sample are indexed from 1 to I, and the siblings in each family are labeled 1 and 2. Sibling 1 is assumed to be older than sibling 2. There are D years under observation, which are labeled from 1 to D. Both members of each sibship i are assumed to be working in all of these years. Let $t_{i,j,d}$ represent the age of sibling j from family i in year d, and let $s_{i,j}$ and $z_{i,j}$ respectively denote the schooling and the test score of sibling j from family i. The age of each person increases by one in each year. Letting $t_{i,0} = (t_{i,1,0}, t_{i,2,0})$ represent the ages of the two siblings from family i in year zero, the set T of possible realizations of $t_{i,0}$ is taken to be finite. Every element of T is assumed to be a pair of distinct nonnegative integers.

Let $b_{i,j}$ be a $K \times 1$ vector of background variables for sibling j from family i. Although these variables were not discussed earlier, there is a simple way to formally introduce them into the framework without changing the predictions of either learning model. Assuming that $b_{i,j}$ is observable both to employers and to the econometrician, let the respective means $\mu_{a,i,j}$, $\mu_{\epsilon,i,j}$, $\mu_{\omega,i,j}$ of $a_{i,j}$, $\epsilon_{i,j}$, $\omega_{i,j}$ have the form:

$$(\mu_{a,i,j}, \mu_{\epsilon,i,j}, \mu_{\omega,i,j}) = \mathbb{E}\left[(a_{i,j}, \epsilon_{i,j}, \omega_{i,j}) | b_{i,1}, b_{i,2}, t_{i,0}\right] = \left(\phi_{a,0} + b_{i,j}'\phi_a,\ \phi_{\epsilon,0} + b_{i,j}'\phi_\epsilon,\ \phi_{\omega,0} + b_{i,j}'\phi_\omega\right), \quad \text{(A.1)}$$

where $\phi_{a,0}$, $\phi_{\epsilon,0}$, and $\phi_{\omega,0}$ are constants, and ϕ_a, ϕ_ϵ, and ϕ_ω are $K \times 1$ coefficient vectors.[1]

Each sibling pair can be represented by the triple (i, p, q), where i indexes the family from which the

[1]The other parameters of the model—β, σ_a^2, ρ_a, γ, σ_ϵ^2, ρ_ϵ, θ_a, θ_s, σ_ω^2, ρ_ω, σ_η^2—are assumed not to depend on the realizations of $b_{i,1}$, $b_{i,2}$, and $t_{i,0}$. The term $h(t_{i,j,d})$ is assumed to be a function only of $t_{i,j,d}$.

two siblings are drawn, and p and q are the respective labels of the first and the second siblings in the pair.[2] I define two vectors:

$$t_{i,(p,q),d} = (t_{i,p,d}, t_{i,q,d})', \quad x_{i,(p,q)} = (z_{i,p}, z_{i,q}, s_{i,p}, s_{i,q}, b'_{i,p}, b'_{i,q})', \tag{A.2}$$

where $t_{i,(p,q),d}$ represents the ages of the two siblings from family i in year d, and $x_{i,(p,q)}$ contains their labor market characteristics. The conditional expectation of the log wage $y_{i,p,d}$ of sibling p from family i in year d given $x_{i,(p,q)}$ and $t_{i,(p,q),d}$ can be put in the following general form both under individual and under social learning:

$$\mathbb{E}\left(y_{i,p,d}|x_{i,(p,q)}, t_{i,(p,q),d}\right) = c(t_{i,(p,q),d}) + x'_{i,(p,q)} v(t_{i,(p,q),d}), \tag{A.3}$$

where $v(t_{i,(p,q),d})$ is a $(2K + 4) \times 1$ coefficient vector, and $c(t_{i,(p,q),d})$ is a constant. Note that $v(t_{i,(p,q),d})$ and $c(t_{i,(p,q),d})$ can vary with the age vector $t_{i,(p,q),d}$ of the two siblings from family i in year d.

I next define the two parameters of interest. For each family i, let G_i be a random variable that takes on the value of each natural number between 1 and D with equal probability $1/D$. Each realization of G_i represents a particular year from the set of observed dates. The random variable G_i is assumed to be independent of all the other variables in the model. Letting $\delta(\tilde{t}_{i,0})$ denote the proportion of families in which the ages of the two siblings in year zero are $\tilde{t}_{i,0} \in T$, the expected value ν_H of $v(t_{i,(1,2),G_i})$ is equal to:

$$\nu_H = \mathbb{E}[v(t_{i,(1,2),G_i})] = D^{-1} \sum_{\tilde{t}_{i,0} \in T} \delta(\tilde{t}_{i,0}) \sum_{d=1}^{D} v(\tilde{t}_{i,(1,2),0} + d\mathbb{1}_2), \tag{A.4}$$

and the expected value ν_L of $v(t_{i,(2,1),G_i})$ is equal to:

$$\nu_L = \mathbb{E}[v(t_{i,(2,1),G_i})] = D^{-1} \sum_{\tilde{t}_{i,0} \in T} \delta(\tilde{t}_{i,0}) \sum_{d=1}^{D} v(\tilde{t}_{i,(2,1),0} + d\mathbb{1}_2), \tag{A.5}$$

where $\mathbb{1}_2$ is a 2×1 vector of ones. For a randomly sampled family, ν_H and ν_L can be interpreted as the average values of the coefficient vectors $v(t_{i,(1,2),G_i})$ and $v(t_{i,(2,1),G_i})$ in a random year.[3]

It is now possible to state the following result. Consider the conditional expectation function in equation (A.3) as well as the expected values of the coefficient vectors in equations (A.4) and (A.5). First, if employer learning is individual, then the ratio of the second to the first entry of ν_H will be equal to the ratio of the second to the first entry of ν_L. That is, under individual learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will be the same as the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage. Second, if employer learning is social, then the ratio of the second to the first entry of ν_H will be less than the ratio of the second to the first entry of ν_L, especially assuming that the first entries of ν_H and ν_L are both positive. That is, under social learning, the ratio of the average coefficient on a younger sibling's test score to the average coefficient on one's own test score in an older sibling's log wage will typically be lower than the ratio of the average coefficient on an older sibling's test score to the average coefficient on one's own test score in a younger sibling's log wage.

[2] Note that each family i contains two sibling pairs: $(i, 1, 2)$ and $(i, 2, 1)$.

[3] Observe that the first and second elements of the vector ν_H (resp. ν_L) represent the average values of the coefficients on one's own and a younger (resp. an older) sibling's test scores in the conditional expectation of an older (resp. a younger) sibling's log wage in equation (A.3).

Proposition A.1 *For $i \in \{1,2\}$, let $\nu_{H,i}$ denote the i^{th} element of the vector ν_H in equation (A.4), and let $\nu_{L,i}$ denote the i^{th} element of the vector ν_L in equation (A.5).*

1. *If learning is individual, then $\nu_{H,2}\nu_{L,1} = \nu_{L,2}\nu_{H,1}$.*

2. *If learning is social, then $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$.*

Proof I begin by proving the first item of the proposition. The parameters $\nu_{H,1}$, $\nu_{H,2}$ and $\nu_{L,1}$, $\nu_{L,2}$ in the statement of the proposition have the following form under individual learning:

$$\begin{aligned} \nu_{H,1} = \mathbb{E}[\chi(t_{i,(1,2),G_i})\pi_o],\ \nu_{H,2} = \mathbb{E}[\chi(t_{i,(1,2),G_i})\pi_f] \\ \nu_{L,1} = \mathbb{E}[\chi(t_{i,(2,1),G_i})\pi_o],\ \nu_{L,2} = \mathbb{E}[\chi(t_{i,(2,1),G_i})\pi_f] \end{aligned}, \tag{A.6}$$

where the constants π_o and π_f are defined in the main appendix. Note that $\chi(t_{i,(p,q),d})$, which varies with only the first element of $t_{i,(p,q),d}$, is the same as the parameter χ_i defined in the main text, where its dependence on t_i was suppressed for ease of notation. Consider the identity:

$$\{\mathbb{E}[\chi(t_{i,(1,2),G_i})]\pi_f\}\{\mathbb{E}[\chi(t_{i,(2,1),G_i})]\pi_o\} = \{\mathbb{E}[\chi(t_{i,(2,1),G_i})]\pi_f\}\{\mathbb{E}[\chi(t_{i,(1,2),G_i})]\pi_o\}. \tag{A.7}$$

Because the constants π_o and π_f can be moved inside each of the expectation signs, it follows from the preceding identity that $\nu_{H,2}\nu_{L,1} = \nu_{L,2}\nu_{H,1}$ as desired.

I next prove the second item of the proposition. The parameters $\nu_{H,1}$, $\nu_{H,2}$ and $\nu_{L,1}$, $\nu_{L,2}$ in the statement of the proposition have the following form under social learning:

$$\begin{aligned} \nu_{H,1} &= \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\pi_f + \xi(t_{i,(1,2),G_i})\pi_o\} \\ \nu_{H,2} &= \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\pi_o + \xi(t_{i,(1,2),G_i})\pi_f\} \\ \nu_{L,1} &= \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\pi_f + \xi(t_{i,(2,1),G_i})\pi_o\} \\ \nu_{L,2} &= \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\pi_o + \xi(t_{i,(2,1),G_i})\pi_f\} \end{aligned}, \tag{A.8}$$

where the constants π_o and π_f are defined in the main appendix. The term $\xi(t_{i,(p,q),d})$, which varies with both the first and second elements of $t_{i,(p,q),d}$, is the same as the parameter ξ_i defined in the main text, where its dependence on t_i and t_e was suppressed for ease of notation. The term $\zeta_r(t_{i,(p,q),d})$, which varies with only the second element of $t_{i,(p,q),d}$, is the same as the parameter ζ_{ri} in the main text, where its dependence on t_e was suppressed for ease of notation.

From the basic properties of the expectation operator, the parameters in equation (A.8) can be rewritten as:

$$\begin{aligned} \nu_{H,1} &= \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_f + \mathbb{E}[\xi(t_{i,(1,2),G_i})]\pi_o \\ \nu_{H,2} &= \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_o + \mathbb{E}[\xi(t_{i,(1,2),G_i})]\pi_f \\ \nu_{L,1} &= \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_f + \mathbb{E}[\xi(t_{i,(2,1),G_i})]\pi_o \\ \nu_{L,2} &= \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_o + \mathbb{E}[\xi(t_{i,(2,1),G_i})]\pi_f \end{aligned}. \tag{A.9}$$

The statement $\nu_{H,2}\nu_{L,1} < \nu_{L,2}\nu_{H,1}$ is equivalent to:

$$\begin{aligned} &\big(\mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_o + \mathbb{E}[\xi(t_{i,(1,2),G_i})]\pi_f\big) \\ &\qquad \cdot \big(\mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_f + \mathbb{E}[\xi(t_{i,(2,1),G_i})]\pi_o\big) \\ <\ &\big(\mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_o + \mathbb{E}[\xi(t_{i,(2,1),G_i})]\pi_f\big) \\ &\qquad \cdot \big(\mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_f + \mathbb{E}[\xi(t_{i,(1,2),G_i})]\pi_o\big) \end{aligned}. \tag{A.10}$$

Expanding both sides of the preceding inequality and canceling out terms appearing on both sides, one obtains after some rearrangement:

$$
\begin{aligned}
&\mathbb{E}[\xi(t_{i,(2,1),G_i})] \cdot \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_o^2 \\
&\qquad + \mathbb{E}[\xi(t_{i,(1,2),G_i})] \cdot \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_f^2 \\
&< \mathbb{E}[\xi(t_{i,(1,2),G_i})] \cdot \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}\pi_o^2 \\
&\qquad + \mathbb{E}[\xi(t_{i,(2,1),G_i})] \cdot \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\}\pi_f^2
\end{aligned}. \tag{A.11}
$$

From the main appendix, the parameters satisfy $1 > \xi_1 > \xi_2 > 0$ and $0 < \zeta_{r1} < \zeta_{r2}$ whenever $t_1 > t_2$. Analogously, we have $1 > \xi(t_{i,(1,2),d}) > \xi(t_{i,(2,1),d}) > 0$ and $0 < \zeta_r(t_{i,(1,2),d}) < \zeta_r(t_{i,(2,1),d})$. It follows that $1 > \mathbb{E}[\xi(t_{i,(1,2),G_i})] > \mathbb{E}[\xi(t_{i,(2,1),G_i})] > 0$ and $0 < \mathbb{E}\{[1-\xi(t_{i,(1,2),G_i})]\zeta_r(t_{i,(1,2),G_i})\} < \mathbb{E}\{[1-\xi(t_{i,(2,1),G_i})]\zeta_r(t_{i,(2,1),G_i})\}$. Thus, equation (A.11) is satisfied if $\pi_o^2 > \pi_f^2$ holds, and it is shown in the main appendix that $\pi_o^2 > \pi_f^2$. ∎

Having shown that the predictions of the employer learning models survive aggregation, I discuss the estimation of the expected values ν_H and ν_L of the coefficient vectors $v(t_{i,(1,2),G_i})$ and $v(t_{i,(2,1),G_i})$. Fixing any nonnegative integer M, let P represent the set composed of every pair of nonnegative integers whose sum is no greater than M. Letting $\#P$ be the number of elements in the set P, the elements of P can be labeled from 1 to $\#P$ with $e^s = (e_1^s, e_2^s)$ denoting the s^{th} element of P. Given a 2×1 vector $t = (t_1, t_2)'$, let f_t denote the $\#P \times 1$ vector whose s^{th} entry is equal to the product $t_1^{e_1^s} t_2^{e_2^s}$; so that, f_t consists of one element for every term of a M^{th}-order bivariate polynomial in t. Let $h_{i,(p,q),d}$ be the $(2K+4+\#P) \times 1$ vector formed by stacking the vector $x_{i,(p,q)}$ on top of the vector $f_{t_{i,(p,q),d}}$. That is, I define:

$$h_{i,(p,q),d} = (x'_{i,(p,q)}, f'_{t_{i,(p,q),d}})', \tag{A.12}$$

where $x_{i,(p,q)}$ comprises the test scores, schooling, and background attributes of the two siblings from family i, and $f_{t_{i,(p,q),d}}$ contains the terms of a bivariate polynomial in their ages in year d.

Some further assumptions become relevant when estimating ν_H and ν_L. Fix $(p,q) = (1,2)$ or $(p,q) = (2,1)$. First, the conditional expectation of $x_{i,(p,q)}$ given that $t_{i,(p,q),G_i} = t$ is assumed to be adequately approximated by a M^{th}-order bivariate polynomial in t. That is, I assume that:

$$\mu_{x,(p,q)}(t) = \sum_{e \in P} \alpha^e_{(p,q)}(t_1^{e_1} t_2^{e_2}), \tag{A.13}$$

where $\mu_{x,(p,q)}(t) = \mathbb{E}(x_{i,(p,q)} | t_{i,(p,q),G_i} = t)$ for any 2×1 vector t of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability, and $\alpha^e_{(p,q)}$ is a $(2K+4) \times 1$ vector that does not depend on t. Second, the matrix representing the expected value of $h_{i,(p,q),G_i} h'_{i,(p,q),G_i}$ is required to be nonsingular. That is, I assume that:

$$\text{rank}\big[\mathbb{E}(h_{i,(p,q),G_i} h'_{i,(p,q),G_i})\big] = 2K + 4 + \#P. \tag{A.14}$$

Third, the variance of $x_{i,(p,q)}$ given that $t_{i,(p,q),G_i} = t$ is restricted to be a matrix of constants that do not vary with t. That is, letting $r_{i,(p,q),G_i} = x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})$, I assume that:

$$\Sigma_{x,(p,q)}(t) = \Sigma_{x,(p,q)}, \tag{A.15}$$

where $\Sigma_{x,(p,q)}(t) = \mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i} | t_{i,(p,q),G_i} = t)$ for any 2×1 vector t of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability, and $\Sigma_{x,(p,q)}$ is a $(2K+4) \times (2K+4)$ matrix of constants that do not depend on t.[4] In addition, note that all random variables are treated as having finite first and second moments.

The following result shows that, under the assumptions above, the parameters ν_H and ν_L can be consistently estimated simply by pooling the observations on each sibling pair across every year and running ordinary least squares regressions on the resulting dataset. In particular, let:

$$\tilde{\nu}_H = \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(1,2),d} h'_{i,(1,2),d} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(1,2),d} y_{i,1,d} \right), \tag{A.16}$$

and let:

$$\tilde{\nu}_L = \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(2,1),d} h'_{i,(2,1),d} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{d=1}^{D} h_{i,(2,1),d} y_{i,2,d} \right). \tag{A.17}$$

Let $\hat{\nu}_H$ and $\hat{\nu}_L$ be vectors containing the first $2K+4$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$, respectively. That is, $\hat{\nu}_H$ (resp. $\hat{\nu}_L$) denotes the estimated coefficient on the covariate vector $x_{i,(1,2)}$ (resp. $x_{i,(2,1)}$) in a log wage regression that also controls for $f_{t_{i,(1,2),d}}$ (resp. $f_{t_{i,(2,1),d}}$). The result below shows that as the number of sampled sibships I goes to infinity, the estimators $\hat{\nu}_H$ and $\hat{\nu}_L$ converge in probability to ν_H and ν_L, respectively.

Proposition A.2 *Suppose that the assumptions in equations (A.13), (A.14), and (A.15) are satisfied. As the number of sampled sibships I goes to infinity, the estimators $\hat{\nu}_H$ and $\hat{\nu}_L$, which consist of the first $2K+4$ elements of $\tilde{\nu}_H$ and $\tilde{\nu}_L$ in equations (A.16) and (A.17), respectively converge in probability to ν_H and ν_L, which are defined in equations (A.4) and (A.5).*

Proof Fix $(p,q) = (1,2)$ or $(p,q) = (2,1)$. The random variable y_{i,p,G_i} can be expressed as:

$$y_{i,p,G_i} = x'_{i,(p,q)} \beta_{(p,q)} + f'_{t_{i,(p,q),G_i}} \gamma_{(p,q)} + e_{i,(p,q),G_i}, \tag{A.18}$$

where $\beta_{(p,q)}$ and $\gamma_{(p,q)}$ are the unique coefficient vectors such that:

$$\mathbb{E}(x_{i,(p,q)} e_{i,(p,q),G_i}) = O_{(2K+4)\times 1} \text{ and } \mathbb{E}(f_{t_{i,(p,q),G_i}} e_{i,(p,q),G_i}) = O_{\#P\times 1}, \tag{A.19}$$

with $O_{(2K+4)\times 1}$ and $O_{\#P\times 1}$ being a $(2K+4)\times 1$ and a $\#P \times 1$ vector of zeros, respectively. Let $\delta_{(p,q)} = (\beta'_{(p,q)}, \gamma'_{(p,q)})'$. Note that $\delta_{(p,q)} = [\mathbb{E}(h_{i,(p,q),G_i} h'_{i,(p,q),G_i})]^{-1} \mathbb{E}(h_{i,(p,q),G_i} y_{i,p,G_i})$ in equation (A.18). Moreover, an alternative expression for y_{i,p,G_i} is:

$$y_{i,p,G_i} = f'_{t_{i,(p,q),G_i}} \theta_{(p,q)} + o_{i,(p,q),G_i}, \tag{A.20}$$

where $\theta_{(p,q)}$ is the unique coefficient vector such that:

$$\mathbb{E}(f_{t_{i,(p,q),G_i}} o_{i,(p,q),G_i}) = O_{\#P\times 1}. \tag{A.21}$$

Note that $\theta_{(p,q)} = [\mathbb{E}(f_{t_{i,(p,q),G_i}} f'_{t_{i,(p,q),G_i}})]^{-1} \mathbb{E}(f_{t_{i,(p,q),G_i}} y_{i,p,G_i})$ in equation (A.20). Finally, the random

[4] This restriction on the conditional variance matrix can be weakened to some extent. Specifically, proposition A.2 remains valid if equation (A.15) is replaced by $\mathbb{E}[\Sigma_{x,(p,q)}(t_{i,(p,q),G_i}) v(t_{i,(p,q),G_i})] = \mathbb{E}[\Sigma_{x,(p,q)}(t_{i,(p,q),G_i})] \mathbb{E}[v(t_{i,(p,q),G_i})]$. That is, the random coefficient vector $v(t_{i,(p,q),G_i})$ is assumed to be uncorrelated with the random conditional variance matrix $\Sigma_{x,(p,q)}(t_{i,(p,q),G_i})$.

vector $x_{i,(p,q)}$ can be decomposed as:

$$x'_{i,(p,q)} = f'_{t_{i,(p,q),G_i}} \lambda_{(p,q)} + u'_{i,(p,q),G_i}, \tag{A.22}$$

where $\lambda_{(p,q)}$ is the unique $\#P \times (2K+4)$ coefficient matrix such that:

$$\mathbb{E}(f_{t_{i,(p,q),G_i}} u'_{i,(p,q),G_i}) = O_{\#P\times(2K+4)}, \tag{A.23}$$

with $O_{\#P\times(2K+4)}$ being a $\#P \times (2K+4)$ matrix of zeros. Note that $\lambda_{(p,q)} = [\mathbb{E}(f_{t_{i,(p,q),G_i}} f'_{t_{i,(p,q),G_i}})]^{-1}\mathbb{E}(f_{t_{i,(p,q),G_i}} x'_{i,(p,q)})$ in equation (A.22). Because the conditional expectation function $\mu_{x,(p,q)}(t_{i,(p,q),G_i})$ in equation (A.13) is assumed to be linear in the elements of $f_{t_{i,(p,q),G_i}}$, one can write $\mu_{x,(p,q)}(t_{i,(p,q),G_i})$ as:

$$[\mu_{x,(p,q)}(t_{i,(p,q),G_i})]' = f'_{t_{i,(p,q),G_i}} \lambda_{(p,q)}, \tag{A.24}$$

where $\lambda_{(p,q)}$ is the same coefficient matrix in equation (A.22) as in equation (A.24).

Now the parameter $\theta_{(p,q)}$ can be expressed as:

$$\begin{aligned}
\theta_{(p,q)} &= [\mathbb{E}(f_{t_{i,(p,q),G_i}} f'_{t_{i,(p,q),G_i}})]^{-1}\mathbb{E}(f_{t_{i,(p,q),G_i}} y_{i,p,G_i}) \\
&= [\mathbb{E}(f_{t_{i,(p,q),G_i}} f'_{t_{i,(p,q),G_i}})]^{-1}\mathbb{E}[f_{t_{i,(p,q),G_i}}(x'_{i,(p,q)}\beta_{(p,q)} + f'_{t_{i,(p,q),G_i}}\gamma_{(p,q)} + e_{i,(p,q),G_i})], \\
&= [\mathbb{E}(f_{t_{i,(p,q),G_i}} f'_{t_{i,(p,q),G_i}})]^{-1}\mathbb{E}(f_{t_{i,(p,q),G_i}} x'_{i,(p,q)})\beta_{(p,q)} + \gamma_{(p,q)} = \lambda_{(p,q)}\beta_{(p,q)} + \gamma_{(p,q)}
\end{aligned} \tag{A.25}$$

where the second step uses equation (A.18) to substitute for y_{i,p,G_i}, and the third step follows from the fact that $\mathbb{E}(f_{t_{i,(p,q),G_i}} e_{i,(p,q),G_i}) = O_{\#P\times 1}$. From equations (A.24) and (A.25), one has:

$$f'_{t_{i,(p,q),G_i}} \theta_{(p,q)} = [\mu_{x,(p,q)}(t_{i,(p,q),G_i})]'\beta_{(p,q)} + f'_{t_{i,(p,q),G_i}}\gamma_{(p,q)}. \tag{A.26}$$

Subtracting equation (A.26) from equation (A.18) yields:

$$y_{i,p,G_i} - f'_{t_{i,(p,q),G_i}}\theta_{(p,q)} = [x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})]'\beta_{(p,q)} + e_{i,(p,q),G_i} = r'_{i,(p,q),G_i}\beta_{(p,q)} + e_{i,(p,q),G_i}. \tag{A.27}$$

Multiplying the left and right sides of the preceding equation by $r_{i,(p,q),G_i}$, one obtains the following after taking the expectation of each side:

$$\begin{aligned}
\mathbb{E}(r_{i,(p,q),G_i} y_{i,p,G_i}) - \mathbb{E}(r_{i,(p,q),G_i} f'_{t_{i,(p,q),G_i}})\theta_{(p,q)} \\
= \mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i})\beta_{(p,q)} + \mathbb{E}(r_{i,(p,q),G_i} e_{i,(p,q),G_i})
\end{aligned}. \tag{A.28}$$

Because $r_{i,(p,q),G_i}$ is by construction orthogonal to any function of $t_{i,(p,q),G_i}$, one has $\mathbb{E}(r_{i,(p,q),G_i} f'_{t_{i,(p,q),G_i}})\theta_{(p,q)} = O_{(2K+4)\times 1}$, noting that $f_{t_{i,(p,q),G_i}}$ is a function of $t_{i,(p,q),G_i}$. In addition, $e_{i,(p,q),G_i}$ is orthogonal to any linear function of $x_{i,(p,q)}$ and $f_{t_{i,(p,q),G_i}}$; so that, $\mathbb{E}(r_{i,(p,q),G_i} e_{i,(p,q),G_i}) = O_{(2K+4)\times 1}$ since $r_{i,(p,q),G_i}$ is linear in $x_{i,(p,q)}$ and $f_{t_{i,(p,q),G_i}}$. Therefore, equation (A.28) implies:

$$\mathbb{E}(r_{i,(p,q),G_i} y_{i,p,G_i}) = \mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i})\beta_{(p,q)}; \tag{A.29}$$

so that, one has:

$$\beta_{(p,q)} = [\mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i})]^{-1}\mathbb{E}(r_{i,(p,q),G_i} y_{i,p,G_i}), \tag{A.30}$$

where the matrix $\mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i})$ is invertible because the matrix $[\mathbb{E}(h_{i,(p,q),G_i} h'_{i,(p,q),G_i})]^{-1}$ is assumed to have full rank as in equation (A.14).

Next, I consider the vector $\mathbb{E}(r_{i,(p,q),G_i} y_{i,p,G_i})$. From equation (A.3), the log wage $y_{i,p,d}$ of sibling p from family i in year d has the following form under both individual and social learning:

$$y_{i,p,d} = c(t_{i,(p,q),d}) + x'_{i,(p,q)} v(t_{i,(p,q),d}) + \varepsilon_{i,(p,q),d}, \tag{A.31}$$

where the error term $\varepsilon_{i,(p,q),d}$ satisfies:

$$\mathbb{E}(\varepsilon_{i,(p,q),d} | x_{i,(p,q)}, t_{i,(p,q),d}) = 0. \tag{A.32}$$

Using equation (A.31), one obtains:

$$\begin{aligned} \mathbb{E}(r_{i,(p,q),G_i} y_{i,p,G_i}) &= \mathbb{E}\{r_{i,(p,q),G_i}[c(t_{i,(p,q),G_i}) + x'_{i,(p,q)} v(t_{i,(p,q),G_i}) + \varepsilon_{i,(p,q),G_i}]\} \\ &= \mathbb{E}[r_{i,(p,q),G_i} c(t_{i,(p,q),G_i})] + \mathbb{E}[r_{i,(p,q),G_i} x'_{i,(p,q)} v(t_{i,(p,q),G_i})] \\ &\quad + \mathbb{E}(r_{i,(p,q),G_i} \varepsilon_{i,(p,q),G_i}) \end{aligned} . \tag{A.33}$$

Let $S_{(p,q)}$ denote the set consisting of every 2×1 vector t of nonnegative integers such that $t_{i,(p,q),G_i} = t$ with positive probability. First, the expectation $\mathbb{E}[r_{i,(p,q),G_i} c(t_{i,(p,q),G_i})]$ can be simplified as follows:

$$\begin{aligned} &\mathbb{E}[r_{i,(p,q),G_i} c(t_{i,(p,q),G_i})] \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t) \mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})] c(t_{i,(p,q),G_i}) | t_{i,(p,q),G_i} = t\} \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t)[\mathbb{E}(x_{i,(p,q)} | t_{i,(p,q),G_i} = t) - \mu_{x,(p,q)}(t)] c(t) = O_{(2K+4)\times 1} \end{aligned} , \tag{A.34}$$

where $O_{(2K+4)\times 1}$ is a $(2K+4) \times 1$ vector of zeros, and $\Pr(t_{i,(p,q),G_i} = t)$ represents the probability that $t_{i,(p,q),G_i} = t$. In equation (A.34), the first equality follows from the law of total expectation and from replacing $r_{i,(p,q),G_i}$ with $x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})$; the second equality follows from the basic properties of the conditional expectation function; and the third equality follows from replacing $\mathbb{E}(x_{i,(p,q)} | t_{i,(p,q),G_i} = t)$ with $\mu_{x,(p,q)}(t)$. Second, the expectation $\mathbb{E}[r_{i,(p,q),G_i} x'_{i,(p,q)} v(t_{i,(p,q),G_i})]$ can be simplified as follows:

$$\begin{aligned} &\mathbb{E}[r_{i,(p,q),G_i} x'_{i,(p,q)} v(t_{i,(p,q),G_i})] \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t) \mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})] x'_{i,(p,q)} v(t_{i,(p,q),G_i}) | t_{i,(p,q),G_i} = t\} \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t) \Big(\mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})] x'_{i,(p,q)} | t_{i,(p,q),G_i} = t\} v(t) \\ &\qquad\qquad - \mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})] \mu'_{x,(p,q)}(t_{i,(p,q),G_i}) | t_{i,(p,q),G_i} = t\} v(t) \Big) \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t) \mathbb{E}(r_{i,(p,q),G_i} r'_{i,(p,q),G_i} | t_{i,(p,q),G_i} = t) v(t) \\ &\quad = \sum_{t \in S_{(p,q)}} \Pr(t_{i,(p,q),G_i} = t) \Sigma_{x,(p,q)} v(t) = \Sigma_{x,(p,q)} \mathbb{E}[v(t_{i,(p,q),G_i})] \end{aligned} . \tag{A.35}$$

In equation (A.35), the first equality follows from the law of total expectation and from substituting

$x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})$ for $r_{i,(p,q),G_i}$; the second equality follows from the basic properties of conditional expectations and from the fact that $\mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})]\mu'_{x,(p,q)}(t_{i,(p,q),G_i})|t_{i,(p,q),G_i} = t\}v(t) = O_{(2K+4)\times 1}$; the third equality follows from the basic properties of conditional expectations and from the definition $r_{i,(p,q),G_i} = x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),G_i})$; the fourth equality follows from the assumption that $\mathbb{E}(r_{i,(p,q),G_i}r'_{i,(p,q),G_i}|t_{i,(p,q),G_i} = t) = \Sigma_{x,(p,q)}$ in equation (A.15); and the fifth equality follows from the law of total expectation. Third, the expectation $\mathbb{E}(r_{i,(p,q),G_i}\varepsilon_{i,(p,q),G_i})$ can be simplified as follows:

$$\begin{aligned}
&\mathbb{E}(r_{i,(p,q),G_i}\varepsilon_{i,(p,q),G_i}) \\
&= D^{-1}\sum_{\tilde{t}_{i,0}\in T}\delta(\tilde{t}_{i,0})\sum_{d=1}^{D}\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|t_{i,0} = \tilde{t}_{i,0}) \\
&= D^{-1}\sum_{\tilde{t}_{i,0}\in T}\delta(\tilde{t}_{i,0})\sum_{d=1}^{D}\mathbb{E}[\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d})|t_{i,0} = \tilde{t}_{i,0}] \\
&= D^{-1}\sum_{\tilde{t}_{i,0}\in T}\delta(\tilde{t}_{i,0})\sum_{d=1}^{D}\mathbb{E}\{[x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),d})]\mathbb{E}(\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d})|t_{i,0} = \tilde{t}_{i,0}\} \\
&= O_{(2K+4)\times 1}
\end{aligned} \quad . \tag{A.36}$$

In equation (A.36), the first and second equalities follow from the law of total expectation; the third equality follows from replacing $r_{i,(p,q),d}$ with $x_{i,(p,q)} - \mu_{x,(p,q)}(t_{i,(p,q),d})$ and from the basic properties of the conditional expectation function; and the fourth equality is due to the fact that $\mathbb{E}(\varepsilon_{i,(p,q),d}|x_{i,(p,q)}, t_{i,(p,q),d})$ $= 0$ by definition. Note that $t_{i,0}$ can be inferred exactly from $t_{i,(p,q),d}$ if the index $[i,(p,q),d]$ is known. To be clear about the notation in equation (A.36), the index $[i,(p,q),G_i]$ is treated as being random when calculating the expectation $\mathbb{E}(r_{i,(p,q),G_i}\varepsilon_{i,(p,q),G_i})$, and the index $[i,(p,q),d]$ is treated as being known when taking the conditional expectation $\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|t_{i,0} = \tilde{t}_{i,0})$. That is, one can also write $\mathbb{E}(r_{i,(p,q),d}\varepsilon_{i,(p,q),d}|t_{i,0} = \tilde{t}_{i,0}) = \mathbb{E}(r_{i,(p,q),G_i}\varepsilon_{i,(p,q),G_i}|G_i = d, t_{i,0} = \tilde{t}_{i,0})$.

Substituting the results from equations (A.34), (A.35), and (A.36) into equation (A.33), one obtains:

$$\mathbb{E}(r_{i,(p,q),G_i}y_{i,p,G_i}) = \Sigma_{x,(p,q)}\mathbb{E}[v(t_{i,(p,q),G_i})]. \tag{A.37}$$

Moreover, it follows from the assumption $\mathbb{E}(r_{i,(p,q),G_i}r'_{i,(p,q),G_i}|t_{i,(p,q),G_i}) = \Sigma_{x,(p,q)}$ in equation (A.15) that $[\mathbb{E}(r_{i,(p,q),G_i}r'_{i,(p,q),G_i})]^{-1} = \Sigma^{-1}_{x,(p,q)}$, where $\Sigma_{x,(p,q)}$ is invertible because $\mathbb{E}(h_{i,(p,q),G_i}h'_{i,(p,q),G_i})$ is assumed to have full rank. Therefore, the parameter $\beta_{(p,q)}$ in equation (A.30) can be expressed as:

$$\beta_{(p,q)} = [\mathbb{E}(r_{i,(p,q),G_i}r'_{i,(p,q),G_i})]^{-1}\mathbb{E}(r_{i,(p,q),G_i}y_{i,p,G_i}) = \mathbb{E}[v(t_{i,(p,q),G_i})] = \nu_J, \tag{A.38}$$

where $\nu_J = \nu_H$ if $(p,q) = (1,2)$ and $\nu_J = \nu_L$ if $(p,q) = (2,1)$. Recall that $\beta_{(p,q)}$ is a $(2K+4)\times 1$ vector that contains the first $(2K+4)$ elements of the full coefficient vector $\delta_{(p,q)} = [\mathbb{E}(h_{i,(p,q),G_i}h'_{i,(p,q),G_i})]^{-1}\mathbb{E}(h_{i,(p,q),G_i}y_{i,p,G_i})$.

Now the estimators $\tilde{\nu}_H$ and $\tilde{\nu}_L$ in equations (A.16) and (A.17) can be expressed as follows. For $\hat{t} \in T$, let $\chi_{i,\hat{t}}$ be an indicator random variable that is equal to one if $t_{i,0} = \hat{t}$ and that is equal to zero otherwise. Letting $J \in \{H, L\}$, one has:

$$\tilde{\nu}_J = (\tilde{\nu}_{J,1})^{-1}\tilde{\nu}_{J,2}, \tag{A.39}$$

where $\tilde{\nu}_{J,1}$ is given by:

$$\tilde{\nu}_{J,1} = D^{-1}\left[I^{-1}\sum_{i=1}^{I}\left(\sum_{\tilde{t}\in T}\sum_{d=1}^{D}\chi_{i,\tilde{t}}h_{i,(p,q),d}h'_{i,(p,q),d}\right)\right], \tag{A.40}$$

and $\tilde{\nu}_{J,2}$ is given by:

$$\tilde{\nu}_{J,2} = D^{-1}\left[I^{-1}\sum_{i=1}^{I}\left(\sum_{\tilde{t}\in T}\sum_{d=1}^{D}\chi_{i,\tilde{t}}h_{i,(p,q),d}y_{i,p,d}\right)\right]. \tag{A.41}$$

Using the weak law of large numbers, one has:

$$\begin{aligned}\text{plim}_{I\to\infty}\, I^{-1}\sum_{i=1}^{I}\left(\sum_{\tilde{t}\in T}\sum_{d=1}^{D}\chi_{i,\tilde{t}}h_{i,(p,q),d}h'_{i,(p,q),d}\right) &= \mathbb{E}\left(\sum_{\tilde{t}\in T}\sum_{d=1}^{D}\chi_{i,\tilde{t}}h_{i,(p,q),d}h'_{i,(p,q),d}\right),\\ = \sum_{\tilde{t}\in T}\sum_{d=1}^{D}\mathbb{E}(\chi_{i,\tilde{t}}h_{i,(p,q),d}h'_{i,(p,q),d}) &= \sum_{\tilde{t}\in T}\sum_{d=1}^{D}\delta(\tilde{t})\mathbb{E}(h_{i,(p,q),d}h'_{i,(p,q),d}|t_{i,0}=\tilde{t})\end{aligned} \tag{A.42}$$

and, using an analogous argument, one has:

$$\text{plim}_{I\to\infty}\, I^{-1}\sum_{i=1}^{I}\left(\sum_{\tilde{t}\in T}\sum_{d=1}^{D}\chi_{i,\tilde{t}}h_{i,(p,q),d}y_{i,p,d}\right) = \sum_{\tilde{t}\in T}\sum_{d=1}^{D}\delta(\tilde{t})\mathbb{E}(h_{i,(p,q),d}y_{i,p,d}|t_{i,0}=\tilde{t}). \tag{A.43}$$

It follows from equations (A.40) and (A.42) that:

$$\text{plim}_{I\to\infty}\,\tilde{\nu}_{J,1} = D^{-1}\sum_{\tilde{t}\in T}\delta(\tilde{t})\sum_{d=1}^{D}\mathbb{E}(h_{i,(p,q),d}h'_{i,(p,q),d}|t_{i,0}=\tilde{t}) = \mathbb{E}(h_{i,(p,q),G_i}h'_{i,(p,q),G_i}), \tag{A.44}$$

and from equations (A.41) and (A.43) that:

$$\text{plim}_{I\to\infty}\,\tilde{\nu}_{J,2} = D^{-1}\sum_{\tilde{t}\in T}\delta(\tilde{t})\sum_{d=1}^{D}\mathbb{E}(h_{i,(p,q),d}y_{i,p,d}|t_{i,0}=\tilde{t}) = \mathbb{E}(h_{i,(p,q),G_i}y_{i,p,G_i}). \tag{A.45}$$

Now, by Slutsky's theorem, equations (A.44) and (A.45) along with equation (A.39) imply that:

$$\text{plim}_{I\to\infty}\,\tilde{\nu}_J = [\mathbb{E}(h_{i,(p,q),G_i}h'_{i,(p,q),G_i})]^{-1}\mathbb{E}(h_{i,(p,q),G_i}y_{i,p,G_i}), \tag{A.46}$$

noting that the matrix $\mathbb{E}(h_{i,(p,q),G_i}h'_{i,(p,q),G_i})$ is assumed to be nonsingular as in equation (A.14). It follows from equation (A.46) that $\text{plim}_{I\to\infty}\,\tilde{\nu}_J = \delta_{(p,q)} = (\beta'_{(p,q)}, \gamma'_{(p,q)})'$, where $\beta_{(p,q)}$ and $\gamma_{(p,q)}$ are the regression parameters appearing in equation (A.18). In addition, recall from equation (A.38) that $\beta_{(p,q)} = \nu_J$. Therefore, as desired, the first $(2K+4)$ elements of $\tilde{\nu}_J$ converge in probability to ν_J. ∎

B Simple Model of Employee Referrals

This appendix develops a simple model of employee referrals that deals with two potential issues. First, the social learning model assumes that one's wage is set equal to the conditional expectation of one's productivity given one's own and a sibling's schooling and performance. If a sibling's characteristics are not observable to a person's employer unless both individuals work for the same firm, then this assumption about wage determination might be unrealistic as a broad description of the labor market. Second, the percentages of individuals obtaining a job through a sibling or also working for the same firm as a sibling are on average moderate in size. If siblings must work for the same firm in order to influence each other's wage, then these percentages might be too small to account for the main estimates of sibling effects.

The model in this section addresses these points by relaxing the assumption that one's employer observes the characteristics of one's sibling and by generating an equilibrium with social effects on wages even if siblings work at different firms. The wage offer made by an informationally advantaged employer is assumed to be observable to other potential employers, who can use this offer to update their beliefs when making counteroffers. In brief, an employer's wage offer may act as a signal to other employers of a worker's productivity.

The basic structure of the model is as follows. There are two siblings and two periods. The siblings differ in seniority with the older and the younger sibling being indexed by 1 and 2, respectively. Each sibling i has a schooling level s_i as well as $B \geq 1$ initial productivity signals $\{r_{iu}\}_{u=1}^{B}$. In period 1, sibling 1 enters the labor market, whereupon each of $M \geq 2$ firms observes s_1 and $\{r_{1u}\}_{u=1}^{B}$. Each of these firms simultaneously makes a wage offer Y_j to sibling 1. Sibling 1 accepts the wage offer of some firm I and works for one period at firm I. Subsequently, firm I observes $C \geq 1$ additional productivity signals $\{r_{1u}\}_{u=B+1}^{B+C}$ for sibling 1. Having worked, sibling 1 refers sibling 2 to firm I and then leaves the labor market.[5] In period 2, sibling 2 enters the labor market, whereupon firm I observes s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$. Firm I makes a wage offer Y_I to sibling 2. Next, $N \geq 2$ other firms observe Y_I as well as s_2 and $\{r_{2u}\}_{u=1}^{B}$. Each of these firms simultaneously makes a wage offer Y_{Oj} to sibling 2, and sibling 2 accepts a wage offer and works for one period. Subsequently, sibling 2's employer observes $C \geq 1$ additional productivity signals $\{r_{2u}\}_{u=B+1}^{B+C}$ for sibling 2.

The additional assumptions of the model are as follows. The properties of the variables s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B+C}$ are as described in the main text. Every wage offer is required to be a positive real number, and each sibling accepts the highest wage offer received.[6] If a firm does not hire a worker in a given period, then the firm obtains a profit of zero for that period. If a firm hires sibling i at wage Y in a given period, then the firm obtains a profit of $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu}) - Y$ for that period, where $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu})$ represents sibling i's output on the job.

The solution concept is perfect Bayesian equilibrium. In period 1, every firm selects Y_j so as to maximize the expected discounted value of its profits given the strategies of the other players as well as its beliefs about each sibling i's output $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu})$ conditional on s_1 and $\{r_{1u}\}_{u=1}^{B}$. In period 2, firm I chooses Y_I so as to maximize the expected value of its profits given the strategies of the other players in addition to its beliefs about sibling 2's output $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u})$ conditional on s_1, s_2 and

[5] It is assumed for simplicity that the older sibling always refers the younger sibling to her employer. The model can be extended to the case where the younger sibling receives a referral from the older sibling with a positive probability less than one. This extension does not change the main prediction of the model, especially if the probability of a referral is independent of the other variables in the model.

[6] In the treatment here, workers are permitted to use mixed strategies when accepting wage offers, although firms are restricted to use pure strategies when making wage offers. The results of the analysis do not change if firms are allowed to randomize over different wage offers.

$\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$. Each remaining employer then chooses Y_{Oj} so as to maximize the expected value of its profits given the strategies of the other players in addition to its beliefs about sibling 2's output $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u})$ conditional on Y_I as well as s_2 and $\{r_{2u}\}_{u=1}^{B}$. Based on the strategies of the players, firms' beliefs are derived from Bayes' rule whenever possible.

In order to solve the model above, I focus on the separating equilibria.[7] The result below establishes the existence of a separating equilibrium. In addition, it shows that in any separating equilibrium, the wage accepted by the older sibling is equal to the conditional expectation of her output given her own schooling and initial productivity signals, and the wage accepted by the younger sibling is equal to the conditional expectation of her output given both siblings' schooling, the younger sibling's initial productivity signals, and all of the older sibling's productivity signals.

Proposition B.1 *There exists a separating perfect Bayesian equilibrium. In any separating equilibrium, the following hold:*

1. *The wage W_1 accepted by sibling 1 is equal to the conditional expectation of* $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{1u})$ *given s_1 and $\{r_{1u}\}_{u=1}^{B}$.*
2. *The wage W_2 accepted by sibling 2 is equal to the conditional expectation of* $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u})$ *given s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$.*

Proof I begin by providing an example of a separating equilibrium. In period 2, after all the wage offers have been made, sibling 2 accepts the wage offer Y_I of firm I if Y_I is greater than the highest wage offer $\max_j Y_{Oj}$ of the other firms. If Y_I is less than or equal to $\max_j Y_{Oj}$, then sibling 2 accepts the wage offer Y_{Ok} of some firm k other than I that makes an offer of $\max_j Y_{Oj}$. If multiple offers by firms other than I are equal to $\max_j Y_{Oj}$, then sibling 2 randomly selects an offer, assigning equal probability to each such offer.

After observing firm I's wage offer Y_I to sibling 2, every other firm believes that $\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}$ is normally distributed with mean $\log(Y_I) - \frac{1}{2}\sigma_I^2$ and variance

$$\sigma_I^2 = \mathbb{V}(\tfrac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}|s_1, s_2, \{r_{1u}\}_{u=1}^{B+C}, \{r_{2u}\}_{u=1}^{B}). \tag{B.1}$$

Each of these firms offers sibling 2 a wage Y_{Oj} equal to Y_I. After observing sibling 1's additional productivity signals $\{r_{1u}\}_{u=B+1}^{B+C}$, firm I believes that $\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}$ is normally distributed with mean

$$\mu_I = \mathbb{E}(\tfrac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}|s_1, s_2, \{r_{1u}\}_{u=1}^{B+C}, \{r_{2u}\}_{u=1}^{B}) \tag{B.2}$$

and variance σ_I^2. Firm I offers sibling 2 a log wage $\log(Y_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$.

In period 1, after observing sibling 1's schooling s_1 and initial productivity signals $\{r_{1u}\}_{u=1}^{B}$, every firm believes that $\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu}$ is normally distributed with mean μ_{Oi} and variance σ_{Oi}^2 where:

$$\mu_{Oi} = \mathbb{E}(\tfrac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu}|s_1, \{r_{1u}\}_{u=1}^{B}) \text{ and } \sigma_{Oi}^2 = \mathbb{V}(\tfrac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu}|s_1, \{r_{1u}\}_{u=1}^{B}). \tag{B.3}$$

Each firm offers sibling 1 a log wage $\log(Y_j)$ equal to $\mu_{O1} + \frac{1}{2}\sigma_{O1}^2$. Sibling 1 accepts the highest wage offer received $\max_j Y_j$. If multiple offers are equal to $\max_j Y_j$, then sibling 1 randomly selects an offer, assigning equal probability to each offer.

[7] To be clear, a separating equilibrium here is a perfect Bayesian equilibrium in which firm I makes a different wage offer Y_I to sibling 2 for each of its possible equilibrium beliefs about sibling 2's output $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u})$ conditional on s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$.

To see that the strategies and beliefs described above form a separating equilibrium, note first that firm I offers a different log wage $\log(Y_I)$ to sibling 2 for each of its possible equilibrium beliefs about $\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}$ conditional on s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$, where any normal distribution with variance σ_I^2 can be an equilibrium belief. Observe next that the specified beliefs are derived from Bayes' rule. In particular, firm I offers sibling 2 a wage $Y_I = \exp(\mu_I + \frac{1}{2}\sigma_I^2)$ equal to the conditional expectation of $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u})$ given s_1, s_2 and $\{r_{1u}\}_{u=1}^{B+C}$, $\{r_{2u}\}_{u=1}^{B}$. Consequently, upon observing Y_I, the other firms use Bayes' rule to infer that $\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{2u}$ is normally distributed with mean μ_I and variance σ_I^2.

It is now straightforward to confirm that the prescribed strategies are sequentially rational given beliefs. Each sibling always accepts the highest wage offer received. In period 2, every firm obtains an equilibrium expected profit of zero. If a firm other than I were to make an offer greater than its equilibrium offer, then it would obtain a negative expected profit. If such a firm were to make an offer less than its equilibrium offer, then it would obtain an expected profit of zero. If firm I were to make an offer different from its equilibrium offer, then it would continue to receive an expected profit of zero, because the other firms would match this offer, and sibling 2 would never choose to work for firm I. In period 1, each firm obtains an equilibrium expected discounted payoff of zero. If a firm were to offer a lower wage, then it would obtain an expected discounted payoff of zero. If a firm were to offer a higher wage, then it would obtain a negative expected discounted payoff, because it would receive a negative expected profit in period 1 and an expected profit of zero in period 2.

I next show that in any separating equilibrium, the accepted wages must be as given in the statement of the proposition. Suppose that a separating equilibrium is being played. First, if firm I offers sibling 2 a log wage $\log(Y_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$, then no other firm k will offer sibling 2 a wage Y_{Ok} greater than or equal to Y_I unless sibling 2 accepts firm k's offer with probability zero. Thus, if firm I offers sibling 2 a log wage $\log(Y_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$, then sibling 2 will accept the offer made by firm I, and firm I will receive a negative expected profit in period 2. However, firm I could obtain an expected payoff of zero in period 2 by instead offering sibling 2 a log wage $\log(Y_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$. Hence, there cannot be a separating equilibrium in which firm I offers sibling 2 a log wage $\log(Y_I)$ greater than $\mu_I + \frac{1}{2}\sigma_I^2$.

Second, if firm I offers sibling 2 a log wage $\log(Y_I)$ equal to $\mu_I + \frac{1}{2}\sigma_I^2$, then no other firm k will make an offer greater than Y_I unless sibling 2 accepts firm k's offer with probability zero. Because sibling 2 always accepts the highest wage offer, it must be in such an equilibrium that no firm offers sibling 2 a log wage greater than $\mu_I + \frac{1}{2}\sigma_I^2$ and that sibling 2 receives a log wage of $\mu_I + \frac{1}{2}\sigma_I^2$. Third, if firm I offers sibling 2 a log wage $\log(Y_I)$ less than $\mu_I + \frac{1}{2}\sigma_I^2$, then there cannot be an equilibrium in which some firm offers sibling 2 a log wage greater than $\mu_I + \frac{1}{2}\sigma_I^2$. Moreover, some firm must offer sibling 2 a log wage equal to $\mu_I + \frac{1}{2}\sigma_I^2$. Otherwise, there would exist a wage offer $\hat{Y}$ greater than $\max(\max_j Y_{Oj}, Y_I)$ but less than $\exp(\mu_I + \frac{1}{2}\sigma_I^2)$ such that some firm k other than I would have an incentive to deviate by offering sibling 2 the wage $\hat{Y}$ instead of making its original wage offer Y_{Ok}. Because sibling 2 always accepts the highest wage offer, it must be in such an equilibrium that sibling 2 receives a log wage of $\mu_I + \frac{1}{2}\sigma_I^2$.

Hence, sibling 2's wage must be as specified in the statement of the proposition. Because every firm obtains an expected profit of zero in period 2, the game played in period 1 is equivalent to Bertrand competition among $M \geq 2$ firms making wage offers to sibling 1, where the total expected output from hiring sibling 1 is equal to the conditional expectation of $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{1u})$ given s_1 and $\{r_{1u}\}_{u=1}^{B}$. Consequently, the highest wage offer made to sibling 1 in such an equilibrium is $\exp(\mu_{O1} + \frac{1}{2}\sigma_{O1}^2)$. Hence, sibling 1's wage must be as specified in the statement of the proposition. ∎

Two remarks should be made in regard to the result above. First, although attention is restricted to the separating equilibria of the model, other equilibria with different implications for wage setting exist. For example, a pooling equilibrium can be constructed in which the wage accepted by each sibling i is equal to the conditional expectation of her total output $\exp(\frac{1}{C}\Sigma_{u=B+1}^{B+C} r_{iu})$ given her own schooling s_i and initial productivity signals $\{r_{iu}\}_{u=1}^{B}$.[8] In such an equilibrium, each sibling's wage depends only on one's own characteristics. Second, in the separating equilibrium described in the first three paragraphs from the proof of proposition B.1, the older sibling's characteristics always affect the younger sibling's log wage, even though the two siblings never work for the same firm. The reason for this outcome is that the wage offer of the older sibling's former employer reveals private information to other firms about the younger sibling's productivity.[9]

The result below shows that if a separating equilibrium is played as in proposition B.1, then the wages of the two siblings have the same basic structure as under the social learning model in the main text. That is, if each sibling's log wage is regressed on both siblings' schooling and test scores, then the ratio of the coefficient on a younger sibling's test score to that on one's own test score in an older sibling's log wage is typically lower than the ratio of the coefficient on an older sibling's test score to that on one's own test score in a younger sibling's log wage. Note that the properties of each sibling's test score z_i are as described in the main text.

Proposition B.2 *Suppose that a separating equilibrium is played as in proposition B.1. Let ϑ_{ij} denote the regression coefficient on sibling j's test score in the conditional expectation of sibling i's log wage given s_1, s_2 and z_1, z_2. Then $\vartheta_{12}\vartheta_{22} < \vartheta_{21}\vartheta_{11}$.*

Proof Under individual learning, the conditional expectation of sibling 1's log wage $\log(W_1)$ given s_1, s_2 and z_1, z_2 has the form:

$$\mathbb{E}[\log(W_1)|s_1, s_2, z_1, z_2] = \chi_1 \mathbb{E}(a_1|s_1, s_2, z_1, z_2) + H_1(s_1), \tag{B.4}$$

where H_1 is some function of s_1, and the parameter χ_1 is defined by:

$$\chi_1 = B\sigma_\eta^{-2}\sigma_{g1}^2, \ \ \sigma_{g1}^2 = (\sigma_m^{-2} + B\sigma_\eta^{-2})^{-1}, \ \ \sigma_m^2 = \mathbb{V}(a_1|s_1). \tag{B.5}$$

Hence, the coefficients ϑ_{11} and ϑ_{12} in the statement of the proposition can be expressed as:

$$\vartheta_{11} = \chi_1\pi_o \ \text{ and } \ \vartheta_{12} = \chi_1\pi_f, \tag{B.6}$$

where π_o and π_f are as defined in the main appendix. Under social learning, the conditional expectation of sibling 2's log wage $\log(W_2)$ given s_1, s_2 and z_1, z_2 has the form:

$$\mathbb{E}[\log(W_2)|s_1, s_2, z_1, z_2] = (1-\xi_2)\zeta_{r2}\mathbb{E}(a_1|s_1, s_2, z_1, z_2) + \xi_2\mathbb{E}(a_2|s_1, s_2, z_1, z_2) + H_2(s_1, s_2), \tag{B.7}$$

where H_2 is some function of s_1 and s_2; ζ_{r2} is equal to $(B + C)$ times the coefficient on r_{1u} in the conditional expectation of a_2 given s_1, s_2, and $\{r_{1u}\}_{u=1}^{B+C}$; and the parameter ξ_2 is defined by:

$$\xi_2 = B\sigma_\eta^{-2}\sigma_{q2}^2, \ \ \sigma_{q2}^2 = (\sigma_{n2}^{-2} + B\sigma_\eta^{-2})^{-1}, \ \ \sigma_{n2}^2 = \mathbb{V}(a_2|s_1, s_2, \{r_{1u}\}_{u=1}^{B+C}). \tag{B.8}$$

Hence, the coefficients ϑ_{21} and ϑ_{22} in the statement of the proposition can be expressed as:

$$\vartheta_{21} = (1-\xi_2)\zeta_{r2}\pi_o + \xi_2\pi_f \ \text{ and } \ \vartheta_{22} = (1-\xi_2)\zeta_{r2}\pi_f + \xi_2\pi_o. \tag{B.9}$$

Note that ζ_{r2} was shown to be positive in the main appendix. Now, the statement $\vartheta_{12}\vartheta_{22} < \vartheta_{21}\vartheta_{11}$ is

[8] In addition, various semi-separating equilibria can be constructed.

[9] Nonetheless, there can also exist a separating equilibrium in which the two siblings always work for the same firm.

equivalent to:

$$(\chi_1\pi_f)\cdot[(1-\xi_2)\zeta_{r2}\pi_f+\xi_2\pi_o]<[(1-\xi_2)\zeta_{r2}\pi_o+\xi_2\pi_f]\cdot(\chi_1\pi_o), \tag{B.10}$$

which reduces to $\pi_f^2<\pi_o^2$. From the main appendix, we have $\pi_o^2>\pi_f^2$, completing the proof. ∎

C Analysis of Antidiscrimination Policies

This appendix constructs a framework to illustrate how social effects in employer learning can impact employment. The model is applied to study group disparities in labor force participation, and government policies to improve equity or efficiency are proposed. There are two periods and two relatives that differ in age. Let 1 and 2 respectively index the older and the younger relative. Race is denoted by $G\in\{B,W\}$, where B signifies the minority group, and W signifies the majority group. Let L_i be the labor productivity of relative $i\in\{1,2\}$. The variables L_1, L_2 are joint normally distributed with common mean $\mu_{L,G}$, identical variance σ_L^2, and correlation ρ_L. The mean productivity of the minority group $\mu_{L,B}$ can differ from the mean productivity of the majority group $\mu_{L,W}$. Let P_B and P_W with $P_B+P_W=1$ be the respective fractions of the population belonging to the minority and majority groups. The reservation value of each individual is R, which represents the payoff to a nonworking person. The labor market is competitive.

Consider first the case where employers do not statistically discriminate based on race G. However, information on relative 1's performance can be used to predict relative 2's productivity and determine relative 2's wage. Assume that $P_B\mu_{L,B}+P_W\mu_{L,W}>R$, which ensures that relative 1 works at the competitive wage. The timeline of events is as follows. In period 1, relative 1 works and is paid a market wage $\widehat{M}_1$ equal to his or her expected productivity. At the end of period 1, employers observe the productivity L_1 of relative 1, and relative 1 leaves the labor market. In period 2, relative 2 decides whether or not to participate in the labor force. The market offers relative 2 a wage $\widehat{M}_2(L_1)$ equal to the conditional expectation of his or her productivity L_2 given the productivity L_1 of relative 1. Relative 2 works if and only if the market wage $\widehat{M}_2(L_1)$ is greater than or equal to the reservation value R. Relative 2 retires at the end of period 2.

The result below shows that younger relatives from a group with lower mean productivity have a smaller employment probability. Employers do not directly discriminate based on racial group. However, older relatives from a less productive group are observed to have worse performance on average, which causes employers to infer that their younger relatives would be less efficient. Consequently, younger relatives from a disadvantaged group are offered a lower market wage and so withdraw from the labor force.

Proposition C.1 *Assume no statistical discrimination based on race G. Let $\Omega_{2,G}$ denote the employment probability of relative 2 from group G. If $\mu_{L,B}<\mu_{L,W}$, then $\Omega_{2,B}<\Omega_{2,W}$. If $\mu_{L,B}>\mu_{L,W}$, then $\Omega_{2,B}>\Omega_{2,W}$.*

Proof The market wage for relative 1 is $\widehat{M}_1=P_B\mu_{L,B}+P_W\mu_{L,W}$, which is greater than R by assumption. The market wage for relative 2 can be calculated as:

$$\begin{aligned}\widehat{M}_2(L_1)=&\mathbb{E}(L_2|L_1)=P_B\mathbb{E}(L_2|L_1,G=B)+P_W\mathbb{E}(L_2|L_1,G=W)\\=&P_B[(1-\rho_L)\mu_{L,B}+\rho_L L_1]+P_W[(1-\rho_L)\mu_{L,W}+\rho_L L_1]\\=&(1-\rho_L)(P_B\mu_{L,B}+P_W\mu_{L,W})+\rho_L L_1\end{aligned}, \tag{C.1}$$

where the second equality follows from the law of total expectation, and the third equality applies the formulas for the conditional distributions of bivariate normal random variables. Since relative 2 works if and only if $\widehat{M}_2(L_1) \geq R$, the employment probability of relative 2 is given by:

$$\begin{aligned}\Omega_{2,G} =& \Pr[\widehat{M}_2(L_1) \geq R] = \Pr[(1-\rho_L)(P_B\mu_{L,B} + P_W\mu_{L,W}) + \rho_L L_1 \geq R] \\ =& 1 - \Phi\{[R - (1-\rho_L)(P_B\mu_{L,B} + P_W\mu_{L,W})]/(\rho_L\sigma_L) - \mu_{L,G}/\sigma_L\}\end{aligned}, \tag{C.2}$$

where Φ denotes the cdf of the standard normal distribution. The preceding expression shows that $\Omega_{2,G}$ is increasing in $\mu_{L,G}$, whence the proposition follows. ∎

Hence, statistical nepotism can generate racial inequalities in market wages and employment rates. The next result shows that policymakers can equalize employment rates between groups by providing an employer subsidy for hiring younger relatives from the less productive group. The same outcome can be achieved with an in-work subsidy to younger relatives from the disadvantaged group.

Proposition C.2 *Assume no statistical discrimination based on race G. Let $\Omega_{2,G}(S)$ denote the employment probability of relative 2 from group G if a subsidy of S is given to an employer for hiring relative 2 from group G or to relative 2 from group G for working. If $\mu_{L,B} < \mu_{L,W}$, then $\Omega_{2,B}[\rho_L(\mu_{L,W} - \mu_{L,B})] = \Omega_{2,W}(0)$. If $\mu_{L,B} > \mu_{L,W}$, then $\Omega_{2,B}(0) = \Omega_{2,W}[\rho_L(\mu_{L,B} - \mu_{L,W})]$.*

Proof Suppose that a subsidy of S is given to an employer for hiring relative 2 from group G or to relative 2 from group G for working. Relative 2 from group G works if and only if $\mathbb{E}(L_2|L_1) + S \geq R$, where the conditional expectation is given by:

$$\mathbb{E}(L_2|L_1) = (1-\rho_L)(P_B\mu_{L,B} + P_W\mu_{L,W}) + \rho_L L_1. \tag{C.3}$$

Hence, the employment probability of relative 2 from group G can be expressed as:

$$\begin{aligned}\Omega_{2,G}(S) =& \Pr[(1-\rho_L)(P_B\mu_{L,B} + P_W\mu_{L,W}) + \rho_L L_1 \geq R - S] \\ =& 1 - \Phi\{[R - S - (1-\rho_L)(P_B\mu_{L,B} + P_W\mu_{L,W})]/(\rho_L\sigma_L) - \mu_{L,G}/\sigma_L\}\end{aligned}, \tag{C.4}$$

where Φ denotes the cdf of the standard normal distribution. It is straightforward to confirm the proposition given the preceding expression. ∎

Consider now the case where employers statistically discriminate based on race G. Moreover, information on relative 1's performance is used to infer relative 2's productivity and decide relative 2's wage. The following is the sequence of actions. In period 1, relative 1 chooses whether or not to join the labor force. The market offers relative 1 a wage $\widetilde{M}_1(G)$ equal to the conditional expectation of his or her productivity L_1 given race G. Relative 1 works if and only if the competitive wage $\widetilde{M}_1(G)$ is greater than or equal to the outside option R. At the end of period 1, employers observe the productivity L_1 of relative 1 if and only if relative 1 was employed, and relative 1 retires.

In period 2, relative 2 chooses whether or not to join the labor force. If relative 1 worked, then the market offers relative 2 a wage $\widetilde{M}_{2,I}(G, L_1)$ equal to the conditional expectation of his or her productivity L_2 given race G and the productivity L_1 of relative 1. In this case, relative 2 works if and only if $\widetilde{M}_{2,I}(G, L_1)$ is no less than R. If relative 1 did not work, then the market offers relative 2 a wage $\widetilde{M}_{2,O}(G)$ equal to the conditional expectation of his or her productivity L_2 given race G. In this case, relative 2 works if and only if $\widetilde{M}_{2,O}(G)$ is no less than R. Relative 2 retires at the end of period 2.

The result below characterizes employment in a competitive equilibrium of the model. The solution is assumed to be noncooperative in that the younger relative cannot make a side payment to the older relative or to a prospective employer. If the mean productivity $\mu_{L,G}$ of group G is less than the reservation value R, then neither relative works. If $\mu_{L,G}$ is no less than R, then the older relative works, and the younger relative decides whether to participate based on how the market wage $\widetilde{M}_{2,I}(G, L_1)$ compares to R.

Proposition C.3 *Assume statistical discrimination based on race G. The competitive outcome is as follows. If $\mu_{L,G} < R$, then neither relative 1 nor relative 2 from group G works. If $\mu_{L,G} \geq R$, then relative 1 from group G works, and relative 2 from group G works if and only if $L_1 \geq [R - (1 - \rho_L)\mu_{L,G}]/\rho_L$.*

Proof Suppose first that $\mu_{L,G} < R$. Relative 1 does not work because the market wage is $\widetilde{M}_1(G) = \mu_{L,G}$, which is less than R. Consequently, relative 2 does not work because the market wage is $\widetilde{M}_{2,O}(G) = \mu_{L,G}$, which is less than R.

Suppose now that $\mu_{L,G} \geq R$. Relative 1 works because the market wage is $\widetilde{M}_1(G) = \mu_{L,G}$, which is no less than R. Relative 2 works if and only if $\widetilde{M}_{2,I}(G, L_1) \geq R$, where the market wage for relative 2 is given by:

$$\widetilde{M}_{2,I}(G, L_1) = \mathbb{E}(L_2|G, L_1) = (1 - \rho_L)\mu_{L,G} + \rho_L L_1. \tag{C.5}$$

The proposition follows after some substitution and rearrangement. ∎

The next question concerns socially efficient employment decisions. For ease of exposition but without loss of generality, assume that there is no discounting between periods. The total product in period $i \in \{1, 2\}$ equals the reservation value R if relative i does not work and equals the productivity L_i of relative i if relative i does work. A Pareto optimum maximizes the conditional expectation of the sum of the total products in periods 1 and 2 given race G. When allocating relative 1 to employment or nonemployment, a social planner does not know the realizations of the productivities L_1 and L_2 of relatives 1 and 2. If relative 1 is employed, then the realization of L_1 but not L_2 is known when selecting the employment status of relative 2. If relative 1 is not employed, then the social planner knows the realization of neither L_1 nor L_2 when assigning relative 2 to a sector.

As the result below shows, the Pareto optimum depends on a cutoff μ_L^*, which is less than R. If the mean productivity $\mu_{L,G}$ of group G is less than μ_L^*, then neither the younger nor the older relative should be employed. If $\mu_{L,G}$ is greater than μ_L^*, then the older relative should be employed, and the younger relative should be assigned a status based on how the conditional expectation $\mathbb{E}(L_2|G, L_1)$ of his or her productivity compares to R.

Proposition C.4 *Assume statistical discrimination based on race G. There exists a threshold $\mu_L^* < R$ such that the socially efficient employment decisions are as follows. If $\mu_{L,G} < \mu_L^*$, then neither relative 1 nor relative 2 from group G works. If $\mu_{L,G} > \mu_L^*$, then relative 1 from group G works, and relative 2 from group G works if and only if $L_1 \geq [R - (1 - \rho_L)\mu_{L,G}]/\rho_L$.*

Proof If neither relative 1 nor relative 2 works, then the conditional expectation of the sum of the total products in periods 1 and 2 given race G is simply $H_0 = 2R$. If relative 1 does not work but relative 2 works, then the conditional expectation of the sum of the total products in periods 1 and 2 given race G is simply $H_1 = R + \mu_{L,G}$.

Consider now the case where relative 1 works and so L_1 is observed when assigning relative 2 to a sector. The conditional expectation of the productivity L_2 of relative 2 given race G and the productivity L_1 of relative 1 is $(1 - \rho_L)\mu_{L,G} + \rho_L L_1$, which is no less than R if and only if $L_1 \geq [R - (1 - \rho_L)\mu_{L,G}]/\rho_L$. Hence, it is socially efficient for relative 2 to work if and only if $L_1 \geq [R - (1 - \rho_L)\mu_{L,G}]/\rho_L$. If relative

2 is efficiently allocated, then the conditional expectation of the sum of the total products in periods 1 and 2 given race G is:

$$H_2 = \mu_{L,G} + R \cdot \Phi\left(\frac{R-\mu_{L,G}}{\rho_L\sigma_L}\right) + \left[\mu_{L,G} + \rho_L\sigma_L\lambda\left(\frac{R-\mu_{L,G}}{\rho_L\sigma_L}\right)\right] \cdot \left[1-\Phi\left(\frac{R-\mu_{L,G}}{\rho_L\sigma_L}\right)\right], \tag{C.6}$$

where $\lambda = \phi/(1-\Phi)$ denotes the inverse Mills ratio with ϕ and Φ being respectively the pdf and cdf of the standard normal distribution. The first term in the preceding expression represents the conditional expectation given G of the total product in period 1, and the second and third terms constitute the conditional expectation given G of the total product in period 2. The second term is the reservation value multiplied by the conditional probability given G that relative 2 is not employed, and the third term is the conditional probability given G that relative 2 works multiplied by the expected productivity of relative 2 conditional on G and the fact that relative 2 works.

Note that $H_2 > H_1$ for $\mu_{L,G} \geq R$ and $H_0 > H_1$ for $\mu_{L,G} < R$, and so it is never socially efficient for relative 1 not to work but for relative 2 to work. It is straightforward to confirm based on the expression above that H_2 is continuous and increasing in $\mu_{L,G}$ with H_2 having a limit of $-\infty$ as $\mu_{L,G}$ approaches $-\infty$ and a limit of ∞ as $\mu_{L,G}$ approaches ∞. In addition, $H_2 > H_0$ for $\mu_{L,G} = R$, where H_0 is constant in $\mu_{L,G}$. It follows using the intermediate value theorem that there exists $\mu_L^* < R$ such that $H_2 < H_0$ for $\mu_{L,G} < \mu_L^*$ and $H_2 > H_0$ for $\mu_{L,G} > \mu_L^*$. The constant μ_L^* is the threshold in the statement of the proposition. ∎

The competitive outcome is not socially efficient if mean productivity $\mu_{L,G}$ is greater than μ_L^* but less than R. In this case, neither relative works under competition, whereas Pareto optimality requires the older relative to work and the younger relative to choose between employment and nonemployment based on the realized productivity of the older relative. The competitive equilibrium is problematic because of insufficient experimentation. The labor force participation of the older relative generates information about the productivity of the younger relative that is useful when assigning the younger relative to a sector. However, the older relative does not account for the positive externality of his or her decision to work.

In principle, one solution might involve Coasian bargaining, whereby the younger relative compensates the older relative for working or reimburses an employer for hiring the older relative. However, liquidity constraints might prevent a younger relative from making the required transfers. The next result shows that policymakers can implement an efficient outcome by subsidizing employers for hiring older relatives. Alternatively, an in-work subsidy to older relatives can correct the market failure.

Proposition C.5 *Assume statistical discrimination based on race G. Suppose that a subsidy $S = R - \mu_{L,G}$ is given to an employer for hiring relative 1 from group G or to relative 1 from group G for working. The market outcome with the subsidy is for relative 1 from group G to work, and relative 2 from group G works if and only if $L_1 \geq [R - (1-\rho_L)\mu_{L,G}]/\rho_L$.*

Proof Suppose that a subsidy S is given to an employer for hiring relative 1 from group G or to relative 1 from group G for working. Relative 1 from group G works if and only if $\mu_{L,G} + S \geq R$. Hence, relative 1 works for $S = R - \mu_{L,G}$. In this case, relative 2 works if and only if $\mathbb{E}(L_2|G, L_1) \geq R$. This condition can be expressed as $(1-\rho_L)\mu_{L,G} + \rho_L L_1 \geq R$ or, equivalently, $L_1 \geq [R-(1-\rho_L)\mu_{L,G}]/\rho_L$. ∎

Note that a subsidy should be provided to older relatives only from a group G whose mean productivity $\mu_{L,G}$ is greater than μ_L^* but less than R. Specifically, if $\mu_{L,B} \in (\mu_L^*, R)$ but $\mu_{L,W} \notin (\mu_L^*, R)$, then an employment subsidy should be provided to the minority but not to the majority because the employment decisions of the majority but not the minority group are efficient in the equilibrium without intervention. Likewise, if $\mu_{L,W} \in (\mu_L^*, R)$ but $\mu_{L,B} \notin (\mu_L^*, R)$, then an employment subsidy should be provided to the majority but not to the minority because the employment decisions of the minority but not the majority group are efficient in the equilibrium without intervention.

Table D.1: Probability of Given Relative Helping Respondent Obtain Most Recent Job

	Entire Sample	Sibship Size 1	2	3	4	5	6	7+
Percentage Receiving Help from:								
Personal Contact	52.35	49.66	51.67	51.70	53.10	51.77	53.66	53.00
Relative	20.05	12.93	17.66	18.64	20.40	21.20	23.41	21.66
Father	5.28	4.08	7.49	5.77	5.71	4.92	4.39	3.68
Mother	3.47	4.08	3.50	4.33	4.10	3.07	3.66	2.03
Brother	2.08	0.00	0.65	1.29	1.72	2.84	3.17	3.57
Sister	2.12	0.00	0.65	1.34	1.94	2.69	3.05	3.63
Uncle	1.15	1.02	1.14	1.03	1.50	0.84	1.46	1.04
Aunt	0.92	0.34	1.30	0.88	1.00	0.84	1.46	0.55
Cousin	1.39	1.36	1.30	0.82	1.39	1.46	1.46	1.98
Percentage Receiving Help from and Working for Same Employer as:								
Personal Contact	35.06	29.93	35.64	33.83	35.03	35.33	36.71	35.90
Relative	14.20	8.16	12.53	12.20	14.19	15.67	16.83	16.22
Father	3.67	2.72	4.88	3.96	3.88	3.53	3.54	2.64
Mother	2.06	2.38	2.20	2.32	2.49	1.69	2.68	1.21
Brother	1.72	0.00	0.57	1.08	1.55	2.38	1.83	3.08
Sister	1.62	0.00	0.65	0.57	1.61	2.07	2.32	3.02
Uncle	0.85	1.02	0.98	0.72	1.00	0.54	1.34	0.71
Aunt	0.67	0.00	0.90	0.72	0.50	0.77	1.10	0.49
Cousin	1.12	1.02	1.06	0.57	1.11	1.31	1.10	1.65

Note: The tabulations include all 9210 individuals in the NLSY79 with non-missing responses to the relevant questions on job search methods in the 1982 survey. Respondents were first asked, "Was there anyone specifically who helped you get your job with [employer name]?" If so, this question was followed by, "Was this person working for [employer name] when you were first offered this job?" Those answering the first question affirmatively were also asked whether this person was a relative and, if so, what was this person's relationship to them.

Table D.2: Impact of Own AFQT and AFQT of Younger or Older Sibling on Joint Work-Wage Outcomes

	Entire Sample					
	Worked		Wage $\geq$ \$5 Also		Wage $\geq$ \$10 Also	
Older Sibling's AFQT $\times$ Younger Sibling	0.0103	0.0121	0.0148	0.0157	0.0110	0.0082
	(0.0053)	(0.0054)	(0.0070)	(0.0070)	(0.0057)	(0.0055)
Younger Sibling's AFQT $\times$ Older Sibling	0.0094	0.0102	-0.0033	-0.0015	-0.0023	-0.0043
	(0.0055)	(0.0052)	(0.0074)	(0.0072)	(0.0059)	(0.0058)
Own AFQT $\times$ Younger Sibling	0.0260	0.0279	0.0615	0.0624	0.0393	0.0366
	(0.0053)	(0.0054)	(0.0074)	(0.0074)	(0.0056)	(0.0057)
Own AFQT $\times$ Older Sibling	0.0356	0.0350	0.0993	0.0982	0.0653	0.0626
	(0.0061)	(0.0062)	(0.0084)	(0.0083)	(0.0072)	(0.0073)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.6546	0.6194	0.0438	0.0476	0.0753	0.1061
Families	2181	2181	2181	2181	2181	2181
Individuals	5195	5195	5195	5195	5195	5195
Sibling Pairs	8032	8032	8032	8032	8032	8032
Observations	123388	123388	123388	123388	123388	123388
	Out of School					
	Worked		Wage $\geq$ \$5 Also		Wage $\geq$ \$10 Also	
Older Sibling's AFQT $\times$ Younger Sibling	0.0146	0.0167	0.0197	0.0202	0.0150	0.0114
	(0.0057)	(0.0057)	(0.0080)	(0.0080)	(0.0067)	(0.0065)
Younger Sibling's AFQT $\times$ Older Sibling	0.0060	0.0072	-0.0018	-0.0008	0.0037	0.0006
	(0.0063)	(0.0059)	(0.0086)	(0.0084)	(0.0071)	(0.0070)
Own AFQT $\times$ Younger Sibling	0.0285	0.0306	0.0794	0.0798	0.0544	0.0506
	(0.0059)	(0.0060)	(0.0088)	(0.0088)	(0.0070)	(0.0071)
Own AFQT $\times$ Older Sibling	0.0374	0.0374	0.1134	0.1120	0.0780	0.0745
	(0.0067)	(0.0067)	(0.0093)	(0.0092)	(0.0082)	(0.0082)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.2301	0.2142	0.0362	0.0384	0.1541	0.1878
Families	2161	2161	2161	2161	2161	2161
Individuals	5149	5149	5149	5149	5149	5149
Sibling Pairs	7948	7948	7948	7948	7948	7948
Observations	104602	104602	104602	104602	104602	104602

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, mother's education, father's education, mother's age, father's age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's outcome is used as the dependent variable for a given pair. The dataset used here is derived by expanding the main estimation sample to include observations on sibling pairs in which one or both members may not have worked since the last interview. In the upper panel, the sample contains observations in which one or both siblings may not yet have left school. In the lower panel, the sample includes only observations in which both siblings have left school for the first time. In the first pair of columns, the dependent variable is an indicator equal to one if the respondent worked since the last interview and equal to zero otherwise. The dependent variable in the second (third) pair of columns is an indicator equal to one if the respondent worked since the last interview at an hourly wage of at least \$5.00 (\$10.00) in 1982-1984 terms and equal to zero otherwise. The p-values from the delta method are reported for the Wald test of the null hypothesis that the coefficient on (Younger Sibling's AFQT $\times$ Older Sibling) times the coefficient on (Own AFQT $\times$ Younger Sibling) is equal to the coefficient on (Older Sibling's AFQT $\times$ Younger Sibling) times the coefficient on (Own AFQT $\times$ Older Sibling).

Table D.3: Impact on Log Wage of Own AFQT and AFQT of Younger or Older Sibling Not Yet Primarily Working

Older Sibling's AFQT × Younger Sibling	0.0056	0.0048	-0.0184	-0.0105
	(0.0245)	(0.0234)	(0.0322)	(0.0280)
Younger Sibling's AFQT × Older Sibling	0.0141	0.0091	-0.0048	-0.0070
	(0.0132)	(0.0134)	(0.0130)	(0.0131)
Own AFQT × Younger Sibling	0.1070	0.0925	0.1084	0.0938
	(0.0269)	(0.0250)	(0.0269)	(0.0249)
Own AFQT × Older Sibling	0.1071	0.0996	0.1115	0.1046
	(0.0157)	(0.0151)	(0.0156)	(0.0152)
Own Schooling	Yes	Yes	Yes	Yes
Sibling's Schooling	No	No	Yes	Yes
Family Background Controls	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.7703	0.8942	0.6771	0.8837
Families / Individuals / Sibling Pairs / Observations	1528 / 2175 / 2670 / 9596			

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, parental education, parental age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage is used as the dependent variable for a given pair. For a given survey year, the sample comprises those individuals in the NLSY79 who have left school for the first time, have non-missing data on their AFQT score and schooling, have a valid wage observation on a full-time job, have non-missing sibling data including birth order and sibship size, and have a non-twin sibling who has not yet spent a year primarily working. An interviewed sibling is classified as primarily working if she has worked in at least half the weeks since the last interview for an average of at least 30 hours per week during the working weeks. In every survey year, any respondent satisfying these criteria is paired with each of her siblings who has not yet spent a year primarily working, and the resulting sample of sibling pairs is divided into two groups based on whether the respondent is older or younger than the inexperienced sibling. The analysis excludes any siblings whose first year spent primarily working cannot be accurately determined because they have a positive number of weeks unaccounted for in the work history data. The p-values from the delta method are reported for the Wald test of the null hypothesis that the coefficient on (Younger Sibling's AFQT × Older Sibling) times the coefficient on (Own AFQT × Younger Sibling) is equal to the coefficient on (Older Sibling's AFQT × Younger Sibling) times the coefficient on (Own AFQT × Older Sibling).

Table D.4: Impact of Own and Sibling's AFQT on Log Wage Immediately Before and After Siblings Reside in Different Geographic Regions

	Entire Sample		Leave for New Job		Stay at Old Job	
Sibling's AFQT $\times$ Before Separated	0.0285	0.0258	0.0245	0.0400	0.0213	0.0203
	(0.0284)	(0.0281)	(0.0486)	(0.0490)	(0.0325)	(0.0355)
Sibling's AFQT $\times$ After Separated	-0.0197	-0.0320	-0.0570	-0.0463	0.0032	0.0141
	(0.0298)	(0.0294)	(0.0516)	(0.0473)	(0.0314)	(0.0331)
Own AFQT $\times$ Before Separated	0.0929	0.0930	0.0792	0.0800	0.0976	0.0951
	(0.0257)	(0.0247)	(0.0399)	(0.0401)	(0.0373)	(0.0359)
Own AFQT $\times$ After Separated	0.1223	0.1060	0.1059	0.0629	0.1171	0.1116
	(0.0277)	(0.0286)	(0.0471)	(0.0481)	(0.0331)	(0.0329)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.0431	0.0351	0.1468	0.1888	0.3937	0.7159
Families	263	263	203	203	220	220
Individuals	598	598	279	279	344	344
Sibling Pairs	692	692	329	329	380	380
Observations	1480	1480	680	680	800	800

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, parental education, parental age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the dependent variable is the log wage observation before or after the siblings are separated. To construct the dataset, the main estimation sample is expanded to include pairs of siblings born in the same year and month as well as sibling pairs in which one or both members may be missing data on their number of older siblings. The resulting sample is used to identify those sibling pairs for which there exists a consecutive pair of survey years such that the two siblings are living in the same Census geographic region of the United States in the first year but not in the second year. The observations on the sibling pair for the first and second years are included in the samples of sibling pairs before and after being separated, respectively. The dataset excludes any sibling pair in which either member is recorded as residing in a region other than one of the four Census geographic regions of the United States. A sibling pair is included in the new-job sample if there is a change between the two years in the CPS job of the sibling whose wage is used as the dependent variable for the pair. Otherwise, the sibling pair is added to the old-job sample. A sibling pair can belong to both the old-job and the new-job samples if the siblings in a family move between regions in multiple survey years. The p-values from the delta method are reported for the Wald test of the null hypothesis that the coefficient on (Sibling's AFQT $\times$ After Separated) times the coefficient on (Own AFQT $\times$ Before Separated) is equal to the coefficient on (Sibling's AFQT $\times$ Before Separated) times the coefficient on (Own AFQT $\times$ After Separated).

Table D.5: Impact on Log Wage of Own AFQT and AFQT of Younger or Older Sibling Working in Same or Different Occupation, Industry, or Geographic Region

Currently Same Region	Currently Same Occupation		Currently Same Industry		Either or Both	
Older Sibling's AFQT × Younger Sibling	0.1043	0.0988	0.1021	0.0916	0.1037	0.0934
	(0.0365)	(0.0333)	(0.0266)	(0.0255)	(0.0259)	(0.0238)
Younger Sibling's AFQT × Older Sibling	-0.0259	-0.0217	-0.0245	-0.0006	-0.0220	-0.0081
	(0.0351)	(0.0338)	(0.0288)	(0.0273)	(0.0264)	(0.0246)
Own AFQT × Younger Sibling	0.0463	0.0417	0.0655	0.0757	0.0525	0.0578
	(0.0344)	(0.0360)	(0.0302)	(0.0291)	(0.0267)	(0.0272)
Own AFQT × Older Sibling	0.2164	0.1990	0.1521	0.1159	0.1759	0.1452
	(0.0282)	(0.0282)	(0.0287)	(0.0277)	(0.0240)	(0.0238)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.0154	0.0129	0.0128	0.0631	0.0035	0.0103
Families	445	445	543	543	693	693
Individuals	970	970	1192	1192	1548	1548
Sibling Pairs	1092	1092	1360	1360	1804	1804
Observations	2204	2204	3718	3718	4900	4900

Currently Different Region	Always Different Occupation		Always Different Industry		Both	
Older Sibling's AFQT × Younger Sibling	-0.0630	-0.0414	0.0052	-0.0013	0.0174	0.0344
	(0.0352)	(0.0357)	(0.0407)	(0.0381)	(0.0489)	(0.0528)
Younger Sibling's AFQT × Older Sibling	0.0003	-0.0070	-0.0481	-0.0618	0.0365	0.0085
	(0.0395)	(0.0407)	(0.0451)	(0.0504)	(0.0557)	(0.0548)
Own AFQT × Younger Sibling	0.0730	0.1046	0.1002	0.1125	0.0325	0.0342
	(0.0413)	(0.0405)	(0.0401)	(0.0376)	(0.0566)	(0.0595)
Own AFQT × Older Sibling	0.0143	0.0144	0.0876	0.0890	-0.0011	-0.0073
	(0.0402)	(0.0438)	(0.0481)	(0.0469)	(0.0624)	(0.0570)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes	Yes	Yes
Family Background Controls	No	Yes	No	Yes	No	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.7903	0.9733	0.3594	0.2932	0.7275	0.8731
Families	245	245	240	240	146	146
Individuals	555	555	545	545	325	325
Sibling Pairs	628	628	618	618	362	362
Observations	2188	2188	2238	2238	1306	1306

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, parental education, parental age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's log wage serves as the dependent variable for a given pair. The four Census geographic regions of the United States are used when determining whether two siblings live in the same or different areas. A pair of siblings is labeled as currently having the same occupation (industry) if they both belong to the same occupation (industry) in the relevant survey year. Two siblings are said to always be in different occupations (industries) if the set of occupations (industries) reported by one sibling is disjoint from the set of occupations (industries) reported by the other sibling over the entire course of the survey. The 2000 Census 3-digit occupation and industry codes are used to classify observations on sibling pairs. Between the 1979 and 2000 rounds of the NLSY79, the occupation and industry of each job were originally recorded as 1970 Census 3-digit codes. These fields are converted to 2000 Census 3-digit codes based on the crosswalks available from the US Census Bureau. The p-values from the delta method are reported for the Wald test of the null hypothesis that the coefficient on (Younger Sibling's AFQT × Older Sibling) times the coefficient on (Own AFQT × Younger Sibling) is equal to the coefficient on (Older Sibling's AFQT × Younger Sibling) times the coefficient on (Own AFQT × Older Sibling).

Table D.6: Relationship of Own AFQT and Height to Schooling and AFQT of Younger or Older Sibling

	AFQT		Height	
Older Sibling's Schooling × Younger Sibling	0.0466 (0.0084)	0.0283 (0.0080)	0.0189 (0.0351)	0.0039 (0.0359)
Younger Sibling's Schooling × Older Sibling	0.0451 (0.0069)	0.0228 (0.0069)	0.0467 (0.0289)	0.0324 (0.0290)
Own Schooling × Younger Sibling	0.1535 (0.0080)	0.1304 (0.0079)	0.0725 (0.0332)	0.0551 (0.0334)
Own Schooling × Older Sibling	0.1544 (0.0082)	0.1348 (0.0079)	0.0429 (0.0313)	0.0275 (0.0326)
Older Sibling's AFQT × Younger Sibling	——	——	0.0948 (0.0945)	0.0709 (0.0925)
Younger Sibling's AFQT × Older Sibling	——	——	0.0886 (0.0789)	0.0785 (0.0801)
Own AFQT × Younger Sibling	——	——	0.3191 (0.0946)	0.2754 (0.0985)
Own AFQT × Older Sibling	——	——	0.2719 (0.0866)	0.2186 (0.0854)
Family Background Controls	No	Yes	No	Yes
Test for equality between sibling schooling coefficients (p-value)	0.8908	0.5946	0.5531	0.5440
Test for equality between own schooling coefficients (p-value)	0.9347	0.6768	0.5150	0.5444
Test for equality between sibling AFQT coefficients (p-value)	——	——	0.9589	0.9500
Test for equality between own AFQT coefficients (p-value)	——	——	0.7021	0.6547
Families / Individuals / Sibling Pairs	1993 / 4726 / 7074			

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable and fixed effects for each of the two siblings' years of birth. Family background controls are indicator variables for sibship size, parental education, parental age, and each of the two siblings' birth orders. The coefficients on all control variables are estimated separately based on whether the respondent is the older or the younger sibling in a given pair. The dataset contains the first observation on every sibling pair in the main estimation sample. However, the third and fourth columns exclude sibling pairs in which either member is missing information on height. The table reports p-values for the Wald tests of the following restrictions: the coefficient on (Older Sibling's Schooling × Younger Sibling) is equal to the coefficient on (Younger Sibling's Schooling × Older Sibling); the coefficient on (Own Schooling × Younger Sibling) is equal to the coefficient on (Own Schooling × Older Sibling); the coefficient on (Older Sibling's AFQT × Younger Sibling) is equal to the coefficient on (Younger Sibling's AFQT × Older Sibling); the coefficient on (Own AFQT × Younger Sibling) is equal to the coefficient on (Own AFQT × Older Sibling).

Table D.7: Impact of Own AFQT and AFQT of Younger or Older Sibling on Non-Wage Outcomes

	Married	Has Kids	Disabled	In Jail
Older Sibling's AFQT × Younger Sibling	0.0046	-0.0186	0.0004	-0.0031
	(0.0090)	(0.0086)	(0.0038)	(0.0014)
Younger Sibling's AFQT × Older Sibling	0.0043	-0.0172	0.0040	-0.0013
	(0.0099)	(0.0097)	(0.0045)	(0.0013)
Own AFQT × Younger Sibling	0.0364	-0.0069	-0.0181	-0.0044
	(0.0106)	(0.0102)	(0.0043)	(0.0016)
Own AFQT × Older Sibling	0.0656	0.0016	-0.0308	-0.0043
	(0.0107)	(0.0114)	(0.0052)	(0.0016)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes
Family Background Controls	No	No	No	No
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.8387	0.5831	0.6579	0.4433
Families / Individuals / Sibling Pairs / Observations	2169 / 5168 / 7988 / 119708			
	Married	Has Kids	Disabled	In Jail
Older Sibling's AFQT × Younger Sibling	0.0074	-0.0168	-0.0005	-0.0033
	(0.0091)	(0.0087)	(0.0038)	(0.0014)
Younger Sibling's AFQT × Older Sibling	0.0095	-0.0166	0.0041	-0.0017
	(0.0098)	(0.0096)	(0.0044)	(0.0013)
Own AFQT × Younger Sibling	0.0404	-0.0049	-0.0186	-0.0047
	(0.0106)	(0.0102)	(0.0042)	(0.0016)
Own AFQT × Older Sibling	0.0662	0.0046	-0.0301	-0.0044
	(0.0108)	(0.0115)	(0.0052)	(0.0015)
Own and Sibling's Schooling	Yes	Yes	Yes	Yes
Family Background Controls	Yes	Yes	Yes	Yes
Test for equality of ratios of own AFQT impact to sibling AFQT impact (p-value)	0.8919	0.5113	0.4849	0.5541
Families / Individuals / Sibling Pairs / Observations	2169 / 5168 / 7988 / 119708			

Note: Huber-White standard errors, clustered at the family level, are reported in parentheses. All specifications control for the race, gender, region of residence, and urban location of the members of each sibling pair. Included also are indicators for missing data on a given variable, a third-order bivariate polynomial in the ages of the two siblings, and a quartic time trend. Family background controls are indicator variables for sibship size, parental education, parental age, and each of the two siblings' birth orders. The coefficients on all control variables, except for the time trend, are estimated separately based on whether the older or the younger sibling's outcome is used as the dependent variable for a given pair. The dataset is constructed by expanding the main estimation sample to include observations on sibling pairs in which one or both members may not have valid wage data on a full-time job and by limiting the resulting sample to observations on sibling pairs in which both members have non-missing data on marital status, presence of children, health restrictions, and residence type. The p-values from the delta method are reported for the Wald test of the null hypothesis that the coefficient on (Younger Sibling's AFQT × Older Sibling) times the coefficient on (Own AFQT × Younger Sibling) is equal to the coefficient on (Older Sibling's AFQT × Younger Sibling) times the coefficient on (Own AFQT × Older Sibling).