LIMITED NETWORK CONNECTIONS AND THE DISTRIBUTION OF WAGES

Kenneth J. Arrow  
Department of Economics  
Stanford University  
arrow@stanford.edu

Ron Borzekowski  
Department of Economics  
Stanford University  
rborz@stanford.edu

July 2000

Preliminary and Incomplete – Please do not cite

Abstract

It is well-known that 50% or more of all jobs are obtained by network connections. As well, statistical studies show that observable individual factors account for only about 50% of the very wide variation in earnings. We link these two facts and present a simple model of network connections, trying to analyze the extent to which it is conceivable that differences in the size of an individual’s network can contribute substantially to inequality. Our results indicate that differences in the number of ties can induce substantial inequality and can explain 15-20% of the unexplained variation in wages. We also show that reasonable differences in the average number of links between blacks and whites can explain the disparity in black and white income distributions.
Motivation

It is well-known that 50% or more of all jobs are obtained by network connections.\(^1\) It is also well known that the proportion is lower for blacks than for whites [references to be supplied]. It is observed that lower wages (and unemployment) are geographically concentrated. It is inferred from this that reference networks play an important role in the labor market. There is an auxiliary hypothesis in this last inference, that geographical propinquity aids network formation and even may be taken as a surrogate for it.

There is another line of research that we wish to connect up with the role of social linkages in job search, namely, the explanation of income inequality, more specifically of inequality in wages. There are of course many obvious determinants of wage inequality. In a typically neoclassical approach, they tend to be supply factors, such as education, native abilities, or family elements in formation of ability to compete in the labor market. These explanations are all real, but there is one problem. Statistical studies show that collectively they account for only about 50% of the variation in earnings (see, most notably, Jenks et. al. [1972]).

According to the first set of results, the opportunities of individual workers are limited by the extent to which they have connections with firms. This fact, by itself, will give rise to income inequality even if all individuals were fundamentally the same, though not the same in all jobs. With limited connections, an individual worker will not find the job for which he or she is best suited. The neoclassical model is one of perfect matching of workers to jobs. Even if, in such a truly competitive equilibrium, everyone would have the same income, inequality may well appear in a model with limited connections due to the randomness of the connections. Further suppose that different individuals have different degrees of connection to the economy. Then, on the average, it would be expected that the most highly connected workers would have the highest incomes (though there might be a variance). Then workers with different numbers of links will on the average have different incomes. Hence, variability in the numbers of links will contribute to income inequality in a way additional to that of the variables which explain fundamental ability to supply goods.

We present a simple model of links and try to analyze the extent to which it is conceivable that differences in the number of links could contribute substantially to inequality. We also make a preliminary stab at examining whether it is conceivable that the differences in the black and white income distributions could be explained largely by average differences in the degree of connection.

Model\(^2\)

This is a static model. The basic idea is the following. Workers have different generalized abilities. A worker’s productivity in any firm is his or her ability times a random factor (log-

---

\(^1\) The classic study is that of Granovetter [1974]. There is an earlier literature going back, according to Granovetter to De Schweinetz [1932], and including Myers and Shultz [1951], Rees [1966], Rees and Shultz [1970], and others.

\(^2\) The model used is a considerable modification of one proposed by B. Knauth [1998], which was intended for the study of unemployment rather than wage levels.
normally distributed). Any given worker has links with a random sample of the firms. Each firm tied to a worker can observe a noisy signal of his or her productivity. Each such firm then calculates the distribution of the worker’s productivity conditional on the signal and the worker’s ability. The firms to which the worker is tied bid for his or her services, using a second-price auction, so that they bid honestly. Under the assumption of risk neutrality, a firm’s bid will be the mean of the posterior distribution of productivity.

Firms not linked to individuals also bid in this model. It is likely that a worker involved in searching for work may send a resume or call an employer at which he or she has no individual link. In such a case, the unlinked firm receives no additional signal above and beyond ‘publicly’ available information; therefore, it uses the prior distribution of ability as the basis for computing the expected value of the worker to it. To be consistent with the second-price auction, we will need to have at least two such uninformed bidders. If a worker has one or zero links to firms, it will therefore be necessary to add two firms unlinked to the individual to the auction.

Of course, there is imperfect matching of firms and workers, as we would expect in actual complex labor markets. We want to examine how the degree of inequality (to be measured by the variance of the logarithm of wages) varies with the number of ties and the accuracy of the productivity signals. Since analytic solutions seem to be infeasible, we use simulation.

*Let us state the model more specifically. There are a large number of workers indexed by \(w\) and a set of firms indexed by \(f\). In the following, we use capital letters to denote natural logarithms of variables denoted by small letters. A worker \(w\) has a productivity, \(p_{wf}\), at firm \(f\). This is determined by the following random process. Worker \(w\) has an ability level, \(a_w\). His or her productivity at any given firm \(f\) is log-normally distributed, \(N(A_w, h_p^{-1})\), where \(h_p\) is the precision (reciprocal of variance) of \(P_{wf}\) conditional on the ability level. Conditional on ability, the worker’s productivity is independently distributed across firms.

Ability is assumed to be common knowledge, evidenced by educational level or other public information. But productivity is at any given firm is not directly observable by either firms or workers. Instead, the tied firms observe a noisy signal of productivity, \(S_{wf} = p_{wf} + u_{wf}\), where \(u_{wf}\) is distributed \(N(0, h_u^{-1})\). Given this signal, each “informed” firm \(f\) can then estimate the expected productivity of worker \(w\) to it.

The number of informed bidders for a single worker is determined by the worker’s social network. Each worker is endowed with a fixed number of links to firms (chosen at random). We also assume that other firms can bid for each worker, but, given their limited information, their bids are always equal to the (publicly known) ability level, \(a_w\), of the worker. This in effect provides a reservation wage and enables the second-price auction to make sense when the worker has 0 or 1 link.

In the results, we find the wage distribution for a fixed ability level. We can find the wage distribution for each number of links and each assumption about the reliability of the signal. Since wage distributions are log-normal, each is characterized by the mean log wage and variance of log wage. To give an interpretation, we also add the mean wage in each group.

To compare with actual income distributions, we cannot use distributions conditional on the number of links. We therefore assume that the links have a Poisson distribution. We can then give the wage distribution for each value of the Poisson parameter and can calibrate on the observed distribution. It is also possible to decompose the variance of log income to find the proportion of variance explained by variation in links.
Finally, we choose our parameters to be the same for blacks and whites, except for the Poisson parameter of the distribution of links, and to be calibrated to the actual distributions of wages. We show that it is possible by suitable choice of parameters to explain the two wage distributions as arising only from a difference in the distribution of links. This connects our discussion with the concept of social capital, originally introduced by G. Loury for this purpose.

### Simulation

The basic mechanics of our simulations involve two steps. In the first, we assume a fixed ability for all workers and a distribution of productivity around this mean. We also choose a value for the precision of the signal and vary the number of ties from 1 to 20. Thus for a given simulation (e.g., signal = .5, ties = 7), the following algorithm is used. First, we generate productivity values and signals for each of the 100,000 individuals at each firm (700,000 values in this example). We then append the two representative firms that bid the baseline ability of the worker. From these bids, we determine the wage of the worker as equal to the second highest bid, and then compute the mean and variance of these log wages across all 100,000 workers. This is done for each specification of signal precision and number of ties.

In the second step, we choose various values to describe the distribution of ties in the population. With our assumption that ties are Poisson distributed, we need only pick one parameter (λ) to describe the distribution. Thus, for each value of λ and value of the signal, we can describe the mean and variance of log income in the population.

### Calibration

As just described, four parameters are needed for each simulation. We require values for the mean and variance of the firm's signal and the mean number of ties in the population. With no empirical data to rely upon, we run all of our simulations at 3 values for the variance of the signal. These are 0, .5, and 1.0.

In order to calibrate the remaining parameters of the model, we rely on some empirical data from the March 1999 Current Population Survey. In 1998, white men who completed high school but had no college education and who worked full time for 50-52 weeks had an average income of approximately $33,123 with a standard deviation of roughly $22,832. For similar black workers, the average wage and standard deviation were $26,223 and $13,687, respectively. These values as well as the mean and variance of log wages are detailed in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Mean Wage</th>
<th>SD Wage</th>
<th>Avg LnWage</th>
<th>Var LnWage</th>
<th>Half Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>33.123</td>
<td>22.832</td>
<td>3.330</td>
<td>0.3875</td>
<td>0.1937</td>
</tr>
<tr>
<td>Black</td>
<td>26.223</td>
<td>13.687</td>
<td>3.121</td>
<td>0.3226</td>
<td>0.1613</td>
</tr>
</tbody>
</table>

As described above, our simulations generate a distribution of income for the population conditional on assumed parameters for the signal variance and the distribution of ties. As such, our method for choosing the values for the distribution of productivity is to ask the following question: Given no signal noise what values for the mean and variance of productivity and for the distribution of ties for Blacks and Whites, will yield a distribution resembling the empirical
distribution above? In other words, can we find reasonable values for the four free parameters that will generate the 'true' wage patterns for these two populations?

*Our only adjustment to this is to target our simulations to only $\frac{1}{2}$ of the variance of lnWages. The motivation for this is the assumption that the observed wages are still based on a heterogeneous population (even after our controls for full-time work and education). Had the data come from individuals with only one fixed ability, the variance should have been smaller. The factor of $\frac{1}{2}$ is arbitrary.

To perform the calibrations, we use a polytope search method over this four dimensional space. *Since we are targeting four values i.e. mean and variance of [log]wages for Blacks and Whites, we inherently have a system of four non-linear equations in four unknowns. From this, we create the distance function equal to the cartesian distance from each of these equations' results to the *targets and we search for any zeros [what are variables?] for this function. Since this search method will converge to any local minima, we grid search over a wide range of reasonable values.

The results of this search yield two solutions to this system of equations. Three other local minima exist, but are not zeros to the distance function. The two solutions are listed in Table 2.

<table>
<thead>
<tr>
<th>Table 2: Calibrated Parameters for Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Solution 1</td>
</tr>
<tr>
<td>Solution 2</td>
</tr>
</tbody>
</table>

**Results**

In this section, we first assess the results of these calibrations and then discuss some of the qualitative features of this model. To begin, note that the mean and variance of productivity values look reasonable. A person paid a wage equal to their expected productivity would earn approximately $15,000 under solution 1 and $21,000 under solution 2. Both of these are far below the average wages described above. Each tie offers the worker the opportunity to find a firm at which they are more productive and therefore the average wage in the population is far above the average productivity. As well, note that in solution 1, the ability level is set much lower relative to solution 2, precisely because of the impact of the additional ties in solution 1.

It still remains to ask whether the number of ties in these simulations is reasonable. While there is little direct evidence on this point, there is some prior empirical work. In 1985, the General Social Survey included questions regarding the social networks of individuals. As Marsden (1987) describes, the “GSS network data concern those persons with whom a respondent ‘discusses important matters’, whether these be family, finances, health, politics, recreation or other things.” As such, it is hard to say that this is a broader or narrower network definition than we would like. Given the inclusion of the word “important”, it is easy to believe that this name generator would lean toward strong ties of a reasonably intimate matter, a view shared by Marsden. However, given any lack of specificity regarding what matters one discusses with these individuals, the list may be overly broad. For now, we use it as a first benchmark in our analysis.
### Table 3: Wage Distribution for Fixed Number of Informed Bidders (plus two uninformed bidders)

<table>
<thead>
<tr>
<th>Signal Variance</th>
<th>Mean Wage</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>8</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15.2374</td>
<td>15.2374</td>
<td>17.3579</td>
<td>20.2634</td>
<td>23.2795</td>
<td>26.1557</td>
<td>33.6228</td>
<td>37.5702</td>
<td>50.6425</td>
<td>51.7114</td>
<td></td>
</tr>
<tr>
<td>Mean Output</td>
<td>15.2374</td>
<td>24.7280</td>
<td>32.0643</td>
<td>38.0205</td>
<td>42.8924</td>
<td>46.9972</td>
<td>56.7705</td>
<td>61.7051</td>
<td>78.0466</td>
<td>79.2981</td>
<td></td>
</tr>
<tr>
<td>MeanLnWage</td>
<td>2.7238</td>
<td>2.7238</td>
<td>2.8204</td>
<td>2.9456</td>
<td>3.0679</td>
<td>3.1791</td>
<td>3.4341</td>
<td>3.5518</td>
<td>3.8670</td>
<td>3.8891</td>
<td></td>
</tr>
<tr>
<td>VarLnWage</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0526</td>
<td>0.1052</td>
<td>0.1397</td>
<td>0.1553</td>
<td>0.1553</td>
<td>0.1439</td>
<td>0.1117</td>
<td>0.1094</td>
<td></td>
</tr>
<tr>
<td>Mean Output</td>
<td>15.2374</td>
<td>18.9906</td>
<td>22.0944</td>
<td>24.6798</td>
<td>26.8627</td>
<td>28.6973</td>
<td>33.1944</td>
<td>35.4816</td>
<td>42.3532</td>
<td>42.9878</td>
<td></td>
</tr>
<tr>
<td>MeanLnWage</td>
<td>2.7238</td>
<td>2.7238</td>
<td>2.7596</td>
<td>2.8105</td>
<td>2.8685</td>
<td>2.9274</td>
<td>3.0822</td>
<td>3.1634</td>
<td>3.3942</td>
<td>3.4120</td>
<td></td>
</tr>
<tr>
<td>VarLnWage</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0141</td>
<td>0.0317</td>
<td>0.0484</td>
<td>0.0614</td>
<td>0.0770</td>
<td>0.0645</td>
<td>0.0630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MeanLnWage</td>
<td>2.7238</td>
<td>2.7238</td>
<td>2.7570</td>
<td>2.8060</td>
<td>2.8582</td>
<td>2.9116</td>
<td>3.0480</td>
<td>3.1167</td>
<td>3.3138</td>
<td>3.3277</td>
<td></td>
</tr>
<tr>
<td>VarLnWage</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0110</td>
<td>0.0253</td>
<td>0.0373</td>
<td>0.0464</td>
<td>0.0566</td>
<td>0.0559</td>
<td>0.0457</td>
<td>0.0443</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Aggregate Wage Distributions by Poisson Parameter (λ)

<table>
<thead>
<tr>
<th>Signal Variance</th>
<th>Mean Ln Wage</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4.8794</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>7.091</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.7621</td>
<td>2.8471</td>
<td>2.9492</td>
<td>3.0534</td>
<td>3.1409</td>
<td>3.1525</td>
<td>3.2433</td>
<td>3.3250</td>
<td>3.3320</td>
<td>3.3978</td>
<td></td>
</tr>
<tr>
<td>Var Ln Wage</td>
<td>0.0247</td>
<td>0.0741</td>
<td>0.1216</td>
<td>0.1564</td>
<td>0.1757</td>
<td>0.1776</td>
<td>0.1876</td>
<td>0.1897</td>
<td>0.1897</td>
<td>0.1871</td>
<td></td>
</tr>
<tr>
<td>Per Var Ties</td>
<td>0.2387</td>
<td>0.2702</td>
<td>0.2774</td>
<td>0.2726</td>
<td>0.2629</td>
<td>0.2614</td>
<td>0.2469</td>
<td>0.2310</td>
<td>0.2295</td>
<td>0.2147</td>
<td></td>
</tr>
<tr>
<td>Mean Ln Wage</td>
<td>2.7387</td>
<td>2.7741</td>
<td>2.8204</td>
<td>2.8713</td>
<td>2.9169</td>
<td>2.9231</td>
<td>2.9736</td>
<td>3.0215</td>
<td>3.0257</td>
<td>3.0661</td>
<td></td>
</tr>
<tr>
<td>Var Ln Wage</td>
<td>0.0065</td>
<td>0.0211</td>
<td>0.0381</td>
<td>0.0538</td>
<td>0.0651</td>
<td>0.0665</td>
<td>0.0757</td>
<td>0.0819</td>
<td>0.0823</td>
<td>0.0855</td>
<td></td>
</tr>
<tr>
<td>Per Var Ties</td>
<td>0.1494</td>
<td>0.1821</td>
<td>0.1972</td>
<td>0.2020</td>
<td>0.2010</td>
<td>0.2006</td>
<td>0.1954</td>
<td>0.1880</td>
<td>0.1872</td>
<td>0.1793</td>
<td></td>
</tr>
<tr>
<td>Mean Ln Wage</td>
<td>2.7377</td>
<td>2.7707</td>
<td>2.8133</td>
<td>2.8596</td>
<td>2.9066</td>
<td>2.9062</td>
<td>2.9512</td>
<td>2.9935</td>
<td>2.9972</td>
<td>3.0325</td>
<td></td>
</tr>
<tr>
<td>Var Ln Wage</td>
<td>0.0052</td>
<td>0.0167</td>
<td>0.0298</td>
<td>0.0415</td>
<td>0.0497</td>
<td>0.0507</td>
<td>0.0571</td>
<td>0.0611</td>
<td>0.0613</td>
<td>0.0631</td>
<td></td>
</tr>
<tr>
<td>Per Var Ties</td>
<td>0.1636</td>
<td>0.1966</td>
<td>0.2106</td>
<td>0.2140</td>
<td>0.2117</td>
<td>0.2111</td>
<td>0.2043</td>
<td>0.1953</td>
<td>0.1945</td>
<td>0.1852</td>
<td></td>
</tr>
</tbody>
</table>
In Marsden’s study, the average size of the network was 3.01, with roughly half of these individuals being kin. These size of the networks do not vary much by gender, but do vary by race and education. Average network size is highest for Whites (3.01) and lowest for Blacks (2.25) and these values are not far from the values we find in solution 2, 3.7 and 2.8 respectively. As well, if one does think that this name generator is too narrow, the values of 7 and 5 ties may also be quite reasonable.

The more specific results of the simulations are included in Table 3 and Table 4, which are both derived from the values in solution 1. Table 3 shows the mean wage, mean output, and the mean and variance of log-wages for various signal variances and level of ties. As expected, the average wage is strictly increasing in the number of ties. This is precisely the result of Montgomery’s (1991) early model of wages and ties. Individuals with 0 or 1 tie earn of $15, 237 (the expectation of their productivity) while a worker with five ties will earn nearly double that amount. Obviously, log-wages are also monotonically increasing.

An interesting pattern emerges however, among the variance of wages. At the lower end, the variance of wages rises as increasing ties offer more and more workers the chance to earn above their productivity. As the number of ties grows however, the variance drops. In the limit, we expect the variance to go to zero as all wages are tending to infinity. This is exactly the pattern we see in Table 3, however it is most apparent in the simulation for a signal variance of 1.0. Here the variance of log wages increases up to eight or nine ties and then falls over the remaining range.

Table 3 also shows us that very little qualitatively differs as the signal variance is altered. With a very noisy signal, the mean wage is lower as employers are forced to place more weight on the publicly available information and less on their individual signals. The variance is also lower for this same reason.

Perhaps the most interesting finding is the disparity between output and wages. In all of these results, the level of output per worker is nearly double their wage. This outcome is the direct result of our assumption that log-wages are normally distributed and the properties of order statistics on normal variables. With low numbers of draws the differences between the first and second order statistics of normal variables are quite large leading to this apparent anomaly. Nevertheless, it does appear that wages themselves are log-normally distributed and do have a large tail to the right, even for relatively homogenous populations.

In Table 4, we detail the mean and variance of log wages for populations indexed by their average number of ties. The two highlighted regions are the results that match our targets. Thus, a population with the specified ability and variance of productivity and with an average of 4.8794 ties would have an average log wage of 3.1409. This is precisely the true average log wage for the Blacks in our sample. As expected these averages of the values in Table 3 exhibit the same properties: mean log wages are increasing in $\lambda$, while the variance of log wages rises and then falls.

Our last main result is included on the line labeled Per Var Ties. Here, we calculate the percentage of overall variance in log wages attributable to differences in the number of ties among individuals. It is rather startling to find that roughly 15% -25% of the variation in log wages is from this source. As described earlier, measurable differences among individuals may explain 50% of the variation in wages. Among the remaining 50%, our model indicates that almost a quarter is attributable to differences in human capital while the remaining three-quarters is the result of variations in the quality of the worker-firm match.
REFERENCES


LOURY, G. [ ]

