The Role of Firms in Gender Earnings Inequality: Evidence from the United States

By Isaac Sorkin

Men are more likely than women to work in both high-wage firms and high-wage industries. This paper documents that the sorting component accounts for 0.09 log points of the 0.33 log point unconditional gender earnings gap in the United States. A core interpretive issue is whether this sorting component reflects discrimination or differences in preferences. The discrimination explanation is that women would like to work at the same firms or in the same industries as men, but are prevented from doing so. The preference explanation is that there are nonpay characteristics that differ across the low- and high-paying firms and women value these nonpay characteristics more than men.

This paper sheds some light on these alternative mechanisms by estimating a search model. The model embeds versions of both the preference and discrimination explanations for why men and women are sorted. The preference explanation is that men and women rank firms differently, and so, given the same set of opportunities, would end up in different firms. The discrimination explanation is that men and women receive a different set of offers, and so, given the same preferences, end up in different firms.

To separate the preference and discrimination explanations, I estimate the model separately by men and women. This procedure assumes that men and women operate in separate labor markets. Estimation gives me separate offer distributions (opportunities), values (preferences), and earnings for men and women firm-by-firm. I then ask in which ways men and women differ, why they are sorted, and what are the consequences of this sorting.

I. Matched Employer-Employee Data

To be able to study the extent to which men and women are sorted into firms requires matched employer-employee data. I use the US Census Bureau’s Longitudinal Employer Household Dynamics (LEHD) dataset. This quarterly dataset is constructed from employer Unemployment Insurance (UI) filings. An important limitation of this dataset is that it does not include hours. I restrict attention to a worker’s annual dominant job, or the job from which they made the most money in a given year. I also impose an earnings threshold. The sample is the same as Sorkin (2016).

II. The Sorting Component of the Gender Earnings Gap

Table 1 shows that the gap in mean annualized log earnings between men and women is 0.335 log points. One explanation for this gap is that men and women work in different firms: 39.8 percent of men’s coworkers are women, while 60.4 percent of women’s coworkers are women. This finding is quantitatively consistent with
the results for the United States in Hellerstein, Neumark, and McInerney (2008).

One explanation for why men and women work in different firms is that the firms that are high-paying for men are not high-paying for women. If high-paying firms for women were different than for men, then based solely on comparative advantage considerations it would not be surprising to see that men and women work in different firms.

To measure the earnings offered by firms, I use the following equation for log earnings (known as the Abowd, Kramarz, and Margolis 1999 decomposition):

\[
\begin{align*}
\log \text{earnings} & = \alpha_w + \Psi_{J(w,t)} + \sum_{x \in \mathcal{X}} x \cdot \beta_{xJw}t + r_{wt}, \\
& \text{where } y_{wt} \text{ is log earnings of person } w \text{ at time } t, \alpha_w \text{ is a person fixed effect, } \Psi_{J(w,t)} \text{ is the firm fixed effect at the employer } j \text{ where worker } w \text{ is employed at time } t \text{ (denoted by } J(w,t)), r \text{ is an error term, and } x \text{ is a set of covariates including higher-order polynomial terms in age.} \\
\end{align*}
\]

By estimating equation (1) for men and women separately, I allow firms that are high-paying for men to be low-paying for women and vice-versa. In addition, because this is identified by workers who switch employers, it removes the time-invariant person effects captured in the \(\alpha_w\), which might include differences in human capital.

To quantify the role of sorting in the earnings gap between men and women, I can use two sets of “prices.” Considering equation (1), the earnings gap is

\[
\sum_{w \in M} \sum_{t} y_{wt} - \sum_{w \in F} \sum_{t} y_{wt},
\]

where \(M\) is the set of male workers, \(F\) is the set of female workers, \(N_m\) is the total number of male person-years and similarly for \(N_f\). I quantify the role of sorting across firms using either the “male” or “female” prices, denoted \(\Psi_f^{m}(w,t)\) and \(\Psi_f^{f}(w,t)\), respectively.

Table 1—The Role of Firms in the Gender Earnings Gap

<table>
<thead>
<tr>
<th></th>
<th>(y)</th>
<th>(\Psi)</th>
<th>(\alpha_{Rosen})</th>
<th>(\Psi + \alpha_{Rosen})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total difference</td>
<td>0.335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men’s “prices”</td>
<td>0.093</td>
<td>-0.089</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Women’s “prices”</td>
<td>0.087</td>
<td>-0.085</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Panel B. Sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men’s “prices”</td>
<td>0.055</td>
<td>-0.056</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Women’s “prices”</td>
<td>0.046</td>
<td>-0.050</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td>Panel C. Offer distribution: overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men’s “prices”</td>
<td>0.063</td>
<td>-0.084</td>
<td>-0.021</td>
<td></td>
</tr>
<tr>
<td>Women’s “prices”</td>
<td>0.065</td>
<td>-0.064</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Panel D. Offer distribution: sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men’s “prices”</td>
<td>0.034</td>
<td>-0.043</td>
<td>-0.009</td>
<td></td>
</tr>
<tr>
<td>Women’s “prices”</td>
<td>0.032</td>
<td>-0.037</td>
<td>-0.005</td>
<td></td>
</tr>
<tr>
<td>Panel E. Correlations between men and women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(V^m)</td>
<td>0.898</td>
<td>0.920</td>
<td>0.867</td>
<td></td>
</tr>
<tr>
<td>Sector</td>
<td>0.980</td>
<td>0.974</td>
<td>0.955</td>
<td></td>
</tr>
</tbody>
</table>

Notes: \(y\) is log annualized earnings, \(\Psi\) is the firm effect in earnings, \(\alpha_{Rosen}\) is the nonpay characteristics, and \(V^m\) is the overall value of being employed at the firm. In panels A–D, the column weights the prices given in the row name by where men are employed, while the women column weights the prices given in the row name by where women are employed. Column 4 sums together columns 2 and 3.

Source: Author’s calculations using the Longitudinal Employer Household Dynamics (LEHD) dataset. See Sorkin (2016) for more details.

One plausible explanation is that the Portuguese data contains hours, and so part of what the sorting component measures is differences in hours; that is, there are high- and low-hours firms, and men are women are sorted on this basis.

2 Because I only use 7 years of data, the linear terms in the age-wage profile are highly correlated with the person fixed effects and following Card, Heining, and Kline (2013) are omitted.
Sorting across sector explains a large share of the sorting component. In panel B of Table 1, I aggregate the firm-specific earnings premia to the sector level (there are 20 of them). The sorting across sectors explains over half of the sorting component.

III. Understanding Sorting

This section discusses several explanations for why men and women are sorted in the labor market and quantifies them. The full model is the same as Sorkin (2016) estimated separately on the subgroups. The key equation in the model is that the value of being employed at firm $i$ is given by the following equation:

$$V_i^e = \omega(\Psi_i + a_{Rosen,i}),$$

where $\Psi$ is the earnings at the firm, and $a_{Rosen}$ are the nonpay aspects of the firm. Both $\Psi$ and $a_{Rosen}$ are in log dollar units. Moreover, all workers search from a common offer distribution.

The model embeds two main explanations for why this sorting might occur. First, preferences might differ. In the model, this is reflected in differences in the gender-specific firm-level values. These differences can arise either because of differences in firm-level pay, or because of differences in (valuations of) nonpay characteristics across genders. Second, opportunities might differ. This is reflected in differences in the gender-specific offer distributions.3

Differences in preferences do not play a large role in explaining sorting. Panel E of Table 1 shows that men and women have quite highly correlated overall values, and valuations of nonpay characteristics at the firm level. At the sector level, the correlations are even higher.

Differences in the offer distribution are likely to explain sorting. Panel C of Table 1 shows that the offer distribution from which women search contains lower-paying firms. The difference is about 0.06 log points.

IV. Consequences of Sorting

Unlike in earnings where it is unambiguous that women are at worse firms than men, once nonpay characteristics are taken into account the answer is less clear. In Table 1, column 3 of panel A shows that women work in firms with more desirable nonpay characteristics than men, whether measured using men’s or women’s prices. These differences in nonpay characteristics almost completely outweigh the difference in pay. Column 4 of panel A combines the differences in pay with the differences in nonpay characteristics. While using men’s prices, men are at firms that pay 0.09 log points more than the firms where women work, once nonpay characteristics are taken into account the gap in overall value is only 0.004 log points. The story with women’s prices is the same. Similarly, column 4 of panel B shows that at the sector level, once nonpay characteristics are taken into account, it is not clear that women are in less desirable sectors than men.

Similarly, panel C and D show that taking into account differences in nonpay characteristics means that the offer distribution that women search from is not necessarily worse than the offer distribution that men search from. Again, this result is robust to using men’s or women’s prices.

V. Discussion

Men and women are sorted across different firms and sectors. This paper uses a random search model to argue that men and women work at different firms because of differences in opportunities, rather than differences in preferences. The paper interprets differences in opportunities as “searching from a different offer distribution.”

The key interpretive issue is where the different offer distribution comes from. In a random search model, the exogenous offer distribution is a reduced-form representation of the complicated process by which workers direct their search toward particular firms, and firms direct their search toward particular workers. Because the offer distribution is a reduced-form representation of this behavior, it might be possible that a richer model of the offer distribution would arrive at a different conclusion than this paper. Nonetheless, it is still instructive and striking that this simple model focuses attention

---

3There is a third explanation as well: it might be harder for women to climb the job ladder. Formally, “search” parameters might differ by gender so that women are more likely to lose their jobs involuntarily, or less likely to receive outside offers.
on the offer distribution, rather than differences in preferences.

REFERENCES


