Measuring Sorting with Time-Varying Characteristics

Robert Shimer
University of Chicago
February 15, 2017

This paper uses a simple assignment model with time-varying characteristics as a laboratory for exploring methods for uncovering evidence of sorting between workers and jobs using matched worker-firm data sets.

I imagine an economy with $N$ workers and $N$ jobs, each of whom is observed for $T$ periods. Each worker is described by a characteristic $x_{i,t} = \bar{x}_i + \varepsilon_{i,t}$, where $\bar{x}_i$ is a random variable with a fixed distribution function and $\varepsilon_{i,t}$ is a random variable with mean zero drawn independently and identically for all $i$ and $t$. Each job is described by a characteristic $y_{j,t} = \bar{y}_j + \eta_{j,t}$, where again $\bar{y}_j$ is a random variable with a fixed distribution function and $\eta_{j,t}$ is a random variable with mean zero drawn independently and identically for all $t$ and $t$.

There is a production function $f(x, y)$ that combines the characteristic of the worker and job to produce some homogeneous output. I assume that $f$ is supermodular and that the labor market is frictionless and competitive in each period. This implies that in every period, the worker $i$ with the $n^{th}$ highest value of $x_{i,t}$ matches with the job $j$ with the $n^{th}$ highest value of $y_{j,t}$ for all $n = 1, \ldots, N$.

With an equal number of workers and jobs, there are generally many wages that support this equilibrium. I look at the highest-wage equilibrium, where the wage of the least productive worker is equal to the output she produces, while the wage of the $n^{th}$ most productive worker is set at a level that leaves the $n^{th}$ most productive job just indifferent between hiring that worker and the $(n+1)^{st}$ most productive worker.

Given this setup, I imagine a data set containing information on the identity of the worker and the job, as well as the wage. We are interested in understanding in recovering from that data set the fixed characteristic of each worker $\bar{x}_i$ and the fixed characteristic of each job $\bar{y}_j$, and possibly also the production function $f$.

Abowd, Kramarz, and Margolis (Econometrica 1999) offers the standard approach to answering this question. They propose regressing

$$\log w_{i,t} = \alpha_i + \beta_{J(i,t)} + u_{i,t},$$

where $\alpha_i$ is a fixed effect for each worker, $\beta_j$ is a fixed effect for each job, and $J(i, t)$ is the job held by worker $i$ in period $t$. Assuming $T \geq 2$, it is feasible to estimate this equation subject to a single normalization, say on the fixed effect of worker 1. The question is whether $\alpha_i$ bears any resemblance to $\bar{x}_i$ and $\beta_j$ to $\bar{y}_j$.

A simple example show the problem with this approach. Consider a large data set, with both big $N$ and big $T$. Suppose $\varepsilon_{i,t} = 0$ for all $i$ and $t$, so workers do not change over time.
On the other hand, the variance of the firm type $\eta_{j,t}$ is strictly positive. Then to the extent that the distribution of firm types $y_{j,t}$ is constant (a large $N$ assumption), the wage $w_{i,t}$ will also be constant over time for each worker. On the other hand, which firm pays that wage will vary over time, so $J(i,t)$ will not be constant. That means that estimating the AKM regression will, roughly speaking, give us that $\alpha_i = \log w_i$, the log of the constant wage paid to that worker, while $\beta_j = u_{i,t} = 0$ for all $i$, $j$, and $t$. While this describes the data well, the conclusions generally do not tell us anything about the parameters of interest.

Of course, if we instead assume that firms’ type is constant and workers’ type varies over time, we get the opposite conclusion: firms’ type determines wages and there is no dispersion in workers’ type.

When I introduce some variance in both components, I find that these points are robust. Call the variance of $\bar{x}_i$ and $\bar{y}_j$ the signal and the variance of $\varepsilon_{i,t}$ and $\eta_{j,t}$ the noise. The type with a higher signal to noise ratio is seen as having more variance in the fixed effect, even if the difference is driven entirely by the noise term.

In this case, I can also examine the correlation between the fixed effects, as a measure of sorting. I find that when $T$ is small, the AKM regression overfits the fixed effects and so generates a spurious negative correlation between the worker and job. As $T$ grows, the bias shrinks but does not seem to disappear.

I am currently exploring alternative estimators that will do a better job of reflecting the true dispersion in $\bar{x}_i$ and $\bar{y}_j$, as well as the correlation between those two outcomes. An example is the estimator proposed by Bonhomme, Lamadon, and Manresa in “A Distributional Framework for Matched Employer Employee Data” and possibly a version of the estimator in Hagedorn, Law, and Manovskii (Econometrica 2017). It is unclear that either approach will solve the problems highlighted here, particularly the first one.