IZA DP No. 3076

Matching Bias in Labor Demand Estimation

Giovanna Aguilar
Silvio Rendon

September 2007
Matching Bias in Labor Demand Estimation

Giovanna Aguilar  
*Universidad Católica del Perú*

Sílvio Rendon  
*Stony Brook University  
and IZA*

Discussion Paper No. 3076  
September 2007

IZA  
P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0  
Fax: +49-228-3894-180  
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.
ABSTRACT

Matching Bias in Labor Demand Estimation*

Using a matched firm-worker dataset, we show both theoretically and empirically that positive assortative matching between firms and workers leads to an underestimation of the absolute value of wage elasticity of labor demand.

JEL Classification: J23, J32

Keywords: matching, employment, labor demand estimation

Corresponding author:

Silvio Rendon
Economics Department
Stony Brook University
NY 11794-0001
USA
E-mail: srendon@itam.mx

* We thank Daniel Hamermesh for his valuable comments and suggestions. All errors and omission are only ours.
1 Introduction

Becker’s (1993) matching criterion has been widely accepted and applied to several economic phenomena such as marriage and labor markets. However, the econometric consequences of assortative matching, for instance for finding unbiased estimates as part of a system of equations, have not been clearly established. In particular, how does the fact that more productive firms hiring more productive workers affect labor demand estimation? In this article, we show both theoretically and empirically that OLS estimation in the presence of positive assortative matching biases negatively, i.e., tends to underestimate wage the absolute value of labor demand elasticities. This issue is crucial, as most studies (see Hamermesh 1993) conducted on this subject find relatively low estimated values for this elasticity.

2 Estimation approach

Let the following a log-linear function represent the long-run unconditional demand for labor

\[ \ln L_i = \beta \ln w_i + X_i \delta + u_i, \]

where \( w_i \) is the wage rate paid by firm \( i \), \( X_i \) are the other components of the demand function observed by the researcher, and \( u_i \) is a random variable with zero mean and variance \( \sigma^2_u \), representing the unobserved components of the demand function for firm \( i \in \{1, 2, ..., N\} \).

If \( \ln w_i \) is uncorrelated with \( u_i \), then one can obtain an unbiased and consistent estimation of \( \beta \) by OLS. However, even if the labor supply is fully elastic and individual (fixed) effects are accounted for, the underlying endogeneity can lead to biased estimated of \( \beta \). To illustrate this assertion, let the infinitely elastic labor supply of individual \( j \in \{1, 2, ..., M_i\} \) related to firm \( i \) be

\[ \ln w_{ij} = Z_{ij} \gamma + v_{ij}, \]
where $w_{ij}$ is the wage, $Z_{ij}$ is a vector of covariates that determine wages, $\gamma$ are its associated parameters, and $v_{ij}$ is a random variable of unobservables with zero mean, variance $\sigma_v^2$ and covariance with $u_i$ equal to $\sigma_{uv}$. This is a Mincer equation in which $Z_{ij}$ only includes supply-side variables, such as education, tenure, experience, and individual workers’ attributes. Then firm $i$ faces a labor supply function:

$$\ln w_i = Z_i \gamma + v_i,$$

where $\ln w_i = M_i^{-1} \sum \ln w_{ij}$, $Z_i = M_i^{-1} \sum Z_{ij}$, and $v_i = M_i^{-1} \sum v_{ij}$ is distributed with mean 0, variance $M_i^{-1}\sigma_v^2$, and covariance with $u_i$ equal to $\sigma_{uv}$.

This is a recursive or limited information estimation model, in which OLS yields unbiased estimates of the labor demand parameters. The crucial assumption for this to be true is that the firm’s individual labor supply is infinitely elastic and error terms are independent, $\sigma_{uv} = 0$. Otherwise, if $\sigma_{uv} > 0$, that is, if unobservables that increase wages are positively correlated with unobservables that increase labor demand, then $\beta$ estimated by OLS will exhibit an upward bias (or, to an underestimation of the wage labor demand elasticity, as $\beta$ is likely to be negative). One can think of this positive correlation as evidence for positive assortative matching between firms and workers: more productive workers are matched to larger firms. On the contrary, if $\sigma_{uv} < 0$, OLS underestimates $\beta$ and there is negative assortative matching between firms and workers.

Figure 1 depicts the supply and demand for labor and illustrates this matching effect: if $D_1$ is matched to $S_1$, $D_2$ to $S_2$, and $D_3$ to $S_3$, then the resulting equilibrium

1Unfortunately, our dataset does not include any variable that could possible proxy workers’ education.

2If $Z_{ij}$ Eq. (2) includes labor demand variables that are excluded in Eq. (1), then it becomes a typical reduced form and one can attempt to identify not only the wage labor demand, but also the wage labor supply elasticity.
points describe a positive relationship or a negative relationship that is steeper than
the labor demand (overestimation of $\beta$); if matching between supply and demand is
the other way round, negative, the relationship resulting from the equilibrium points
is flatter than the labor demand (underestimation of $\beta$).

Under this setup we can propose the following estimation procedure:

1. First Stage: Estimate Eq. (2), predict workers’ wages, and aggregate them to
   $$\ln w_i = Z_i \hat{\gamma}.$$ 

2. Second Stage: Estimate Eq. (1) using the predicted $\hat{\ln w_i}$ as a regressor.

This estimation procedure removes the correlation between wages paid for by
employers and the disturbance term in Eq. (1) and yields unbiased estimates of $\beta$.\(^3\)

3 Data

This estimation requires firm-level data matched with data on invididuals working at
the firm. Our data come from the Wage and Salary National Survey (ENSYS) for
June 2004 carried out by the Ministry of Labor of Peru. This is a biquarterly survey
applied in June and December, which comprises private firms of 10 and more workers
and is representative for the main cities (Metropolitan Lima and urban areas of 24
main cities in the country), economic sectors and activities, and firms sizes in Peru.

The survey contains a section with aggregate firm-level information such as the
total number of workers, wages by occupational category, total hours worked. It
also includes a section with information on a sample of individual workers inside
the firm, with variables such as age and gender of the worker, hours worked, basic
wage or salary, legal workers’ deductions and employers’ contributions, and other
nonpermanent payments. Ths survey consists of 1,772 firms, for which there are

\(^3\)An augmented version of this procedure can be applied to measure the employment and dead-
weight loss effects of non-wage labor costs. See Aguilar and Rendon (2007).
19,770 workers of which we select white-collar workers, a subsample of 1714 firms and 13097 workers,

Table 1 shows the descriptive statistics for the variables in the final sample by section of the survey. For several variables there are both firm-level as well as individual information; and understandably, there is more dispersion for the latter than for the former. Hours of work typically exceed 40 hours a week. Around 70% of firms and 77% individuals work in the service sector. Most firms are small, more than 50% employ 50 workers or less; however, more than 50% of workers work in firms with 100 or more workers. Around half of the firms are located in Lima City, the capital of the country. Unionization rates are between 8 and 14%, and females represent 38% of workers. On average workers are around 38 years old and have around 6 years of tenure.

4 Estimation results

In Table 2 we present the wage elasticities of demand for labor measured both by total hours of work and number of workers. In the first column we report an OLS estimation using reported firm-level average wages, in the second column an OLS estimation using an average wage constructed from the individual information, and in the third column an estimation that accounts for endogeneity.

Explanatory variables besides wages are indicators of location, presence of a union in the firm, and sector of activity. These variables capture differences in capital prices
across regions, labor relations across firms, and technology across industrial sectors. Exogenous sources of variation for endogeneity correction are individuals’ age, tenure, and gender which determine workers’ productivity.

Notice that there is no big difference in estimating by OLS the wage elasticity with the reported or the constructed firm average wage. The sign in both estimations is wrong and only becomes negative once endogeneity is corrected for. Thus, for both measures of firms’ employment, by hours of work or number of workers, an estimation that accounts for endogeneity (IV) yields a larger wage elasticity of labor demand than one that is done by OLS, which suggests the existence of positive assortative matching between firms and workers. More productive workers are matched with larger firms.

5 Conclusions

Using a matched firm-workers dataset we have shown that an estimation that accounts for endogeneity of wages yields a larger labor cost elasticity of a long run, non-conditional labor demand than one obtained by OLS. We explain that this result is evidence for positive assortative matching between firms and workers: larger firms are matched with more productive workers.

References


### Table 1. Descriptive Statistics for Firm and Worker Samples. Standard errors in small fonts

<table>
<thead>
<tr>
<th>Survey’s Section</th>
<th>Firm</th>
<th>Worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours of work</td>
<td>44.5</td>
<td>46.1</td>
</tr>
<tr>
<td>Employment</td>
<td>99.8</td>
<td></td>
</tr>
<tr>
<td>Wages</td>
<td>542.2</td>
<td>605.0</td>
</tr>
<tr>
<td>Economic Sector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>6.4</td>
<td>4.8</td>
</tr>
<tr>
<td>Industry</td>
<td>23.4</td>
<td>18.6</td>
</tr>
<tr>
<td>Services</td>
<td>70.2</td>
<td>76.6</td>
</tr>
<tr>
<td>Lima Met.</td>
<td>48.6</td>
<td>55.6</td>
</tr>
<tr>
<td>Union</td>
<td>8.2</td>
<td>13.6</td>
</tr>
<tr>
<td>Women</td>
<td></td>
<td>37.4</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>37.8</td>
</tr>
<tr>
<td>Tenure</td>
<td></td>
<td>6.1</td>
</tr>
<tr>
<td>Nobs.</td>
<td>1714</td>
<td>13097</td>
</tr>
</tbody>
</table>

### Table 2. Estimated employment wage-elasticities measured as total hours and number of workers. Standard errors in small fonts

<table>
<thead>
<tr>
<th></th>
<th>Total hours</th>
<th>Number of workers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ln w_m$</td>
<td>$\ln w$</td>
</tr>
<tr>
<td>$\ln w_m$</td>
<td>0.1827</td>
<td>0.1683</td>
</tr>
<tr>
<td></td>
<td>0.0519</td>
<td>0.0532</td>
</tr>
<tr>
<td>Union</td>
<td>1.2340</td>
<td>1.2499</td>
</tr>
<tr>
<td></td>
<td>0.1233</td>
<td>0.1225</td>
</tr>
<tr>
<td>Lima</td>
<td>0.5628</td>
<td>0.5799</td>
</tr>
<tr>
<td></td>
<td>0.0817</td>
<td>0.0808</td>
</tr>
<tr>
<td>Const.</td>
<td>5.7021</td>
<td>5.7535</td>
</tr>
<tr>
<td></td>
<td>0.1760</td>
<td>0.1750</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.179</td>
<td>0.178</td>
</tr>
<tr>
<td>Nobs</td>
<td>1714</td>
<td>1714</td>
</tr>
</tbody>
</table>

$\ln w_m$: Log of the Average Firm-level Wage (Firms’ sample)  
$\ln w$: Firm-level Average of Log-Wage (Workers’ sample)  
$\ln w_p$: Average Firm-level predicted Log-Wage (Workers’ sample)
Figure 1: Labor Supply and Demand