

Matching Bias in Labor Demand Estimation*

Giovanna Aguilar^a, Sílvia Rendon^b

^a*Departamento de Economía, Universidad Católica del Perú, Lima 32, Perú*

^b*Economics Department, Stony Brook University, NY 11794, USA*

September 2007

Abstract.- 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.

Keywords: Matching, Employment, Labor Demand Estimation.

JEL Classification: J23, J32.

*Our emails: gaguila@pucp.edu.pe, 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

$$\ln L_i = \beta \ln w_i + X_i \delta + u_i, \quad (1)$$

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 σ_u^2 , representing the unobserved components of the demand function for firm $i \in \{1, 2, \dots, N\}$.

If $\ln w_i$ is uncorrelated with u_i , then one can obtain an unbiased and consistent estimation of β 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 β . To illustrate this assertion, let the infinitely elastic labor supply of individual $j \in \{1, 2, \dots, M_i\}$ related to firm i be

$$\ln w_{ij} = Z_{ij} \gamma + v_{ij}, \quad (2)$$

where w_{ij} is the wage, Z_{ij} is a vector of covariates that determine wages, γ are its associated parameters, and v_{ij} is a random variable of unobservables with zero mean, variance σ_v^2 and covariance with u_i equal to σ_{uv} . This is a Mincer equation in which Z_{ij} only includes supply-side variables, such as education,¹ tenure, experience, and individual workers' attributes.² Then firm i faces a labor supply function:

$$\ln w_i = Z_i \gamma + v_i, \quad (3)$$

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 σ_{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 β estimated by OLS will exhibit an upward bias (or, to an underestimation of the wage labor demand elasticity, as β 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 β and there is *negative assortative matching* between firms and workers.

[Figure 1 here]

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 β); 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 β).

Under this setup we can propose the following estimation procedure:

1. First Stage: Estimate Eq. (2), predict workers' wages, and aggregate them to $\widehat{\ln w_i} = Z_i \widehat{\gamma}$.
2. Second Stage: Estimate Eq. (1) using the predicted $\widehat{\ln w_i}$ 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 β .³

3 Data

This estimation requires firm-level data matched with data on individuals 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. This survey consists of 1,772 firms, for which there are

³An 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 here]

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.

[Table 2 here]

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

- Aguilar, G. and Rendon, S. (2007), Employment and Deadweight Loss Effects of Observed Non-Wage Labor Costs. IZA Discussion Paper No. 2856.
- Becker, G. (1993), *A Treatise on the Family*, Harvard University Press, Cambridge.
- Hamermesh, D. (1986), The Demand for Labor in the Long Run, *in* O. Ashenfelter and E. R. Layard, eds, 'Handbook of Labor. Economics, Volume', North Holland., Amsterdam, pp. 53-90.

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

Survey's Section	Firm	Worker
Hours of work	44.5	46.1
Employment	99.8	
Wages	542.2	605.0
Economic Sector		
Primary	6.4	4.8
Industry	23.4	18.6
Services	70.2	76.6
Lima Met.	48.6	55.6
Union	8.2	13.6
Women		37.4
Age		37.8
Tenure		6.1
Nobs.	1714	13097

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

	Total hours			Number of workers		
	$\ln w_m$	$\overline{\ln w}$	$\overline{\ln w_p}$	$\ln w_m$	$\overline{\ln w}$	$\overline{\ln w_p}$
β	0.1827	0.1683	-0.7825	0.3258	0.2282	-0.6636
	0.0519	0.0532	0.1929	0.0519	0.0521	0.1967
Union	1.2340	1.2499	1.6876	1.2076	1.2615	1.6583
	0.1233	0.1225	0.1497	0.1225	0.1216	0.1472
Lima	0.5628	0.5799	1.1762	0.5270	0.5895	1.1258
	0.0817	0.0808	0.1400	0.0779	0.0791	0.1353
Const.	5.7021	5.7535	7.4431	0.3565	1.8253	5.9748
	0.1760	0.1750	0.3659	0.3297	0.1720	1.1121
R^2	0.179	0.178	0.182	0.213	0.202	0.199
Nobs		1714			1714	

$\ln w_m$: Log of the Average Firm-level Wage (Firms' sample)

$\overline{\ln w}$: Firm-level Average of Log-Wage (Workers' sample)

$\overline{\ln w_p}$: Average Firm-level predicted Log-Wage (Workers' sample)

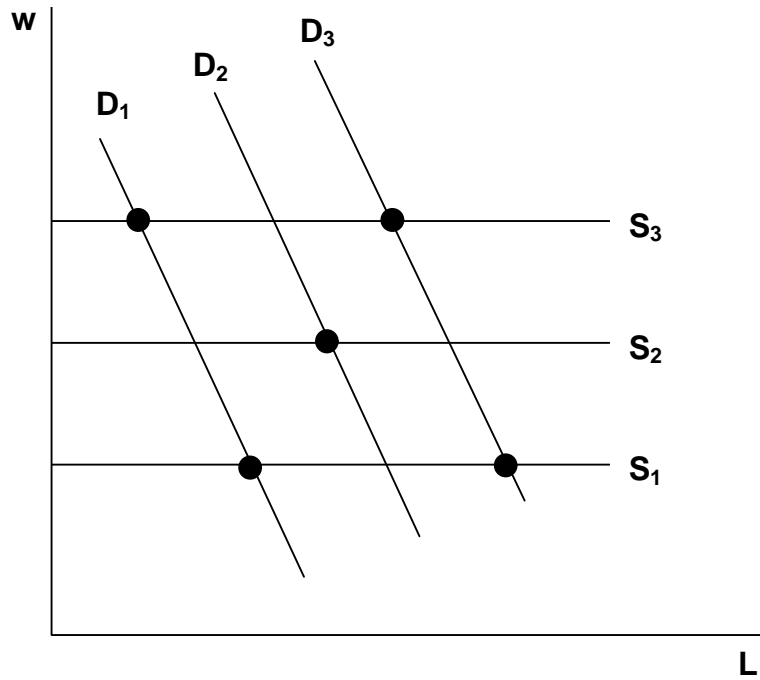


Figure 1: Labor Supply and Demand