4

The Elasticity of the Labor Supply Curve to an Individual Firm

THE single most important idea in this book is that the wage elasticity of the labor supply curve (ε_{Nw} in the notation of previous chapters) is not infinite or close to it. Hence, the most direct way to establish the existence of employer market power over its workers is to estimate the wage elasticity of the labor supply curve facing the firm. Studies of this elasticity are few and far between: one might cite Reynolds (1946a), Nelson (1973), Sullivan (1989), Machin et al. (1993), Boal (1995), Beck et al. (1998), Staiger et al. (1999), and Falch (2001) as an almost complete list. This lack of literature contrasts with entire books written about the demand for labor or the supply of labor by individuals and with the literature on industrial organization on estimating the extent of product market power (for a survey, see Bresnahan 1989). It is testament to the faith that most labor economists have that $\varepsilon_{Nw} = \infty$. But, given the paucity of the literature, this is nothing but faith and some of us might want some evidence that ε_{Nw} is "high" if not infinite.

The plan of this chapter is as follows. The first section discusses the problems of using correlations between wages and employment to estimate the wage elasticity of the labor supply curve facing the firm. We review the literature on the employer size–wage effect, arguing (in the third section) that the evidence suggests that part (though not all) of the employer–size wage effect is the result of an upward-sloping labor supply curve to the individual firm. We argue that, in the absence of good instruments in the form of firm-level demand shocks, it is hard to get a good estimate in this way of the wage elasticity of the labor supply curve facing an individual employer. The fourth section then discusses an alternative method based on using a dynamic monopsony model and estimating the elasticity of separations and recruits with respect to the wage. Finally, the chapter discusses estimates derived from more structural estimation of equilibrium search models.

The main conclusion is that the elasticity of the labor supply curve facing the firm does not seem to be close to infinite but that it is hard to get a very precise estimate of it. An estimate in the region of 2–5 seems to be reasonable. These estimates imply that employers have sizeable amounts of monopsony power: even an elasticity of 5 implies that wages will (using (2.3)) be 17% below the marginal revenue product.

4.1 The Employer Size–Wage Effect

A simple-minded approach to estimating the elasticity of the labor supply curve facing the firm would be to simply regress the log of the wage that the firm pays on the log of employment plus any other variables that might be thought to be important controls. If the market is perfectly competitive we should find a coefficient of zero on employment whereas monopsony would predict it to be positive and the size of the coefficient would give an estimate of the extent of the monopsony power possessed by the firm (as it estimates $\varepsilon = (1/\varepsilon_{Nw})$). There is a large empirical literature that estimates this type of regression and finds a significant positive relationship between wages and employment—what is commonly known as the employer size–wage effect (ESWE). However, an upward-sloping labor supply curve is not the only explanation proposed for the ESWE (for surveys, see Brown and Medoff 1989; Brown et al. 1990; or Oi and Idson 1999) and plausible alternatives need to be considered. Indeed, much of the literature on the ESWE does not even consider an upward-sloping labor supply curve to an individual employer as a possible explanation.[1] This is in spite of the fact that an innocent might think that the first hypothesis an economist would investigate when observing a positive relationship between a price (the wage) and a quantity (employment) is that it is a supply curve. However, Brown and Medoff (1989: 1056) do not manage to identify the cause of the employer size–wage effect and conclude that "our analysis leaves us uncomfortably unable to explain it." Here, we argue that monopsony can fill that void.

Consider a very simple stripped-down model. Assume that firm i has a revenue function which is given by

$$Y_i = \frac{1}{1-\eta} A_i N_i^{1-\eta} \tag{4.1}$$

where A_i is a shock to the marginal revenue product of labor (MPRL) curve. On the supply side of the labor market, assume that the wage that the firm pays is given by

$$w_i = B_i N_i^{\varepsilon} \tag{4.2}$$

[1] For example, the otherwise very thorough and even-handed survey of the relationship between employer size and wages of Brown and Medoff (1989) does not mention monopsony at all with the possible exception that there is some discussion of the rather involved and convoluted "labor pools" model of Weiss and Landau (1984) which could be seen as a sort of monopsony model.

where B_i is a shock to the supply curve. These supply shocks could represent differences in local labor market conditions (because of skill or regional differences) or differences in the attractiveness of non-wage attributes in different firms. We are interested in obtaining a consistent estimate of ε, the inverse elasticity of the labor supply curve facing the firm.

The firm will choose a level of employment where the MPRL equals the marginal cost of labor so that the chosen employment level will satisfy

$$A_i N_i^{-\eta} = (1 + \varepsilon) B_i N_i^{\varepsilon} \tag{4.3}$$

or, in log-linear form

$$\log(N_i) = \frac{1}{\varepsilon + \eta}\left[a_i - b_i - \ln(1 + \varepsilon)\right] \tag{4.4}$$

where $a = log(A)$ and $b = log(B)$. The chosen wage will be given by

$$\log(w_i) = \frac{1}{\varepsilon + \eta}\left[\varepsilon a_i + \eta b_i - \varepsilon \ln(1 + \varepsilon)\right] \tag{4.5}$$

(4.4) and (4.5) are easy to understand. Positive shocks to the MRPL cause employment and wages to rise, although there is only an effect on wages to the extent that the employers do have some labor market power ($\varepsilon > 0$). Positive shocks to the labor supply curve cause employment to fall and wages to rise.[2]

Now make the following assumptions about the observability of the shocks (a,b):

$$a_i = \beta_a x_i + \nu_{ai}$$

$$b_i = \beta_b x_i + \nu_{bi} \tag{4.6}$$

where x is a set of explanatory variables observable to the researcher. In the interests of notational simplicity assume that the same variables affect both a and b. Of course, a particular variable can be constrained to affect only demand or supply shocks by imposing the restriction that its coefficient in the other equation is zero. Assume that the shocks ν are independent of x and jointly normally distributed with mean zero and covariance matrix Σ. Denote by σ_a^2 the variance of ν_a, σ_b^2 the variance of ν_b and σ_{ab} the covariance between ν_a and ν_b.

Now consider how one might set about estimating ε. First, one might think about estimating by OLS the relationship between the log wage and

[2] The model used here is simple in the sense that it uses a static labor supply curve and assumes the employer can only use the wage to influence the supply of labor to the firm: Appendix 4B presents analyses of dynamic labor supply curves and what happens if the generalized model of monopsony of section 2.3 is used.

log employment controlling for other factors thought to be relevant (the x variables in our notation). If one thought of the aim of this as being to estimate a supply curve facing the firm, then one might think of including only those x variables which affect supply (i.e., exclude those affecting only demand). However, many researchers have not thought of their purpose as estimating the supply curve facing a firm so have not used this argument for excluding some variables. For example, when estimating earnings functions, it is common practice to include employment as simply another regressor and the researcher does not think of their aim as being to estimate a supply curve to the individual firm. So, again, we would like to have some idea of the consequences of using this "kitchen sink" approach or using the regressions of others to estimate ε.

A regression of $\log(w_i)$ on $(\log(N_i), x_i)$ estimates $E(\log(w_i) \mid \log(N_i), x_i)$. The following proposition tells us what we would expect to find.

Proposition 4.1. *Running a regression of* $\log(w_i)$ *on* $(\log(N_i), x_i)$ *estimates*

$$E\big(\log(w_i) \mid \log(N_i), x_i\big) = \big(\varepsilon + \rho(\varepsilon + \eta)\big)\log(N_i)$$
$$+\rho\ln(1+\varepsilon) + (\beta_b - \rho(\beta_a - \beta_b))x_i \tag{4.7}$$

where

$$\rho \equiv \frac{\sigma_{ab} - \sigma_b^2}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{ab}} \tag{4.8}$$

Proof. See Appendix 4A.

(4.7) says that the kitchen sink approach will only give an unbiased estimate of ε if $\rho = 0$ which implies that ν_a can be written as ν_b plus some uncorrelated noise. A special case of this is when there are no unobserved supply shocks. In this case, all firms have the same labor supply curve (conditional on x) and variation in N caused by unobserved demand shocks will trace out the labor supply curve. In any other situation one will end up with a biased estimate of ε. If ν_a and ν_b are uncorrelated, the estimate of ε has a downward bias as (4.8) then implies that $\rho < 0$. Intuitively unobserved shifts in the labor supply curve cause wages and employment to move in opposite directions making the slope of the supply curve seem less positive than it really is.

(4.7) can also be used to understand the arguments that the estimated employer size–wage effect overstates the true value of ε (which must be the case if one believes that labor markets are competitive and $\varepsilon = 0$). The two main arguments are unobserved labor quality and compensating wage differentials. One would expect high-quality workers to have a high level of a (as their productivity is high) and a high level of b as b will partly reflect the wages paid by other firms. So, one would expect unobserved labor quality to result in $\sigma_{ab} > 0$. But, from (4.7) and (4.8) this is not sufficient to imply an upward bias: for that we require $\sigma_{ab} > \sigma_b^2$ (or, equivalently that the expectation of $(a - b)$ is increasing in b). If, for example proportional differences in a are reflected in proportional differences in b, then this is exactly the situation in which we obtain an unbiased estimate of ε. However, the fact that workers with high levels of observable skills are more likely to work in large firms does suggest that the condition $\sigma_{ab} > \sigma_b^2$ might be satisfied if the same is true of unobserved skills.[3] One can also understand compensating differentials as a positive correlation between a and b (as the disamenity must have some positive effect on productivity) so that the previous discussion is also relevant for this case. Let us consider the evidence that all of the ESWE can be explained through these effects.

4.2 Competing Explanations for the Employer Size–Wage Effect

Among the strongest contenders for an explanation of the size–wage effect (apart from an upward-sloping labor supply curve) are

- unobserved worker quality;
- compensating wage differentials;
- rent sharing.

All of these possibilities are discussed by Brown and Medoff (1989) in their survey and much of the discussion here is similar.

Table 4.1 presents some basic information on the size–wage effect for the United States (from the April 1993 Contingent Worker Survey (CWS) supplement to the CPS) and the United Kingdom (from the LFS). In both countries, information on employer size is banded, the bands used differing slightly in the two countries. The measure of employer size used is workplace size although the CPS also has information on firm size (which also seems to have a positive impact on wages independent of workplace size). The distribution of workers by establishment size reported in the column headed "sample percentages" is remarkably similar in both countries, the median worker being in a workplace with slightly more than 50 workers.

The column marked (1) for both countries presents estimates of the size–wage effect when there are no other controls in a wage equation. The

[3] In the United States, 12.6% of college graduates work in plants with less than 10 employees and 34% in plants with more than 250. For those who are not college graduates, the figures are 26% in plants with less than 10 workers and 24% in those with more than 250. The United Kingdom is similar.

TABLE 4.1
The Employer Size–Wage Effect in the United States and the United Kingdom

Number of Employees	*United States (April 1993 CPS)*			*United Kingdom (LFS)*		
	Sample Percentage	*(1)*	*(2)*	*Sample Percentage*	*(1)*	*(2)*
1–10	19.9	−0.226 (0.022)	−0.118 (0.019)	18.4	−0.268 (0.004)	−0.151 (0.004)
11–19	13.5	−0.044 (0.024)	−0.015 (0.020)	9.5	−0.097 (0.005)	−0.040 (0.004)
20–24				4.4	−0.048 (0.007)	−0.018 (0.005)
25–49	14.9	0	0	12.5	0	0
50–99	13	0.098 (0.024)	0.067 (0.020)			
100–249	13.1	0.163 (0.024)	0.098 (0.020)	55.2	0.157 (0.006)	0.073 (0.006)
250+	27.6	0.289 (0.020)	0.182 (0.018)			
Other controls	n.a.	No	Yes	n.a.	No	Yes
Number of observations	7854	7854	7854	220868	220868	220868
R^2	n.a.	0.1	0.36	n.a.	0.07	0.42

Notes.
1. The dependent variable is the log of the hourly wage. The other controls included are marital status, children, experience, tenure (all interacted with gender), region, race, and (for the LFS) month dummies. Standard errors are reported in parentheses.

reference category is a workplace with 25–49 workers. The gap in average log wages between the smallest (1–10 employees in both the CPS and the LFS) and the largest workplaces (250+ employees in the United States, 50+ in the United Kingdom) is 0.515 log points in the United States and 0.425 in the United Kingdom. Introducing controls (the columns marked (2)), reduces the magnitude of the effect to approximately half. In both countries it is the introduction of controls for education, experience, and tenure that has the biggest effect in reducing the size–wage effect.

The reduction in the size–wage effect when controls are introduced suggests that part of the raw size–wage differential can be explained by differences in worker quality. But, are estimates that control for worker quality inevitably better than those that do not? To answer this question, let us return to (4.7).

Controlling for labor quality is likely to reduce the unobserved parts of the labor supply and MRPL equations (ν_a and ν_b in (4.6)) so that σ_a^2 and σ_b^2 are reduced in magnitude as, presumably, is σ_{ab}. For a single equation, the standard formula for the extent of omitted variable bias might lead one to believe that the bias is reduced and the resulting estimates are "better." But, matters are more subtle in a simultaneous equations model as the unexplained part of the regressor (here, employer size) is also reduced by the introduction of controls. In fact, (4.7) and (4.8) show that it is (σ_a^2/σ_b^2) and the correlation coefficient between ν_a and ν_b that determines the bias so it is relative, not absolute, variances that are important in determining whether the introduction of controls reduces or increases the bias.

Why does the coefficient on employer size tend to fall when other controls are included? If one thinks that it is mostly "labor quality" and regional variables that induce the correlation between ν_a and ν_b, then we might expect that the correlation between these residuals falls when we improve our controls for these variables. If (for want of a better reason) we assume the relative variances are constant then, using (4.8), it is simple to check that a fall in the correlation between ν_a and ν_b reduces the coefficient on employer size. But, is this reduced coefficient a better estimate of the true value of ε? Not necessarily: as the previous discussion has made clear, reducing the correlation between ν_a and ν_b to zero will lead to an underestimate of the true labor supply elasticity (set $\sigma_{ab} = 0$ in (4.8)). Hence, one should not leap to the conclusion that controlling for labor quality inevitably leads to a better estimate of ε.

However, in spite of this, the reduction in the size–wage effect when controls are introduced does suggest that part of the raw size–wage differential might be explained by differences in worker quality. As a large part of worker quality is unobserved, this has led some to argue that all of the size–wage effect might be explained by differences in worker quality. A common way of controlling for unobserved worker quality is to use panel data to estimate a fixed-effects model: that is, regress changes in wages on changes in employer size. Neither the CPS nor the LFS used above are panel data sets so we use the BHPS for the United Kingdom to investigate this. The results are reported in table 4.2. To make the presentation of the results simple, estimates of a simple elasticity of wages with respect to employer size are presented. Log employer size is computed using the mid-points of the reported bands: more sophisticated attempts to predict employer size conditional on characteristics made little difference to the results.

The first four rows present estimates of the elasticity from the CPS and LFS both with and without controls. Without controls the elasticity of wages with respect to employer size is 0.11 in the United States and 0.14

TABLE 4.2
The Elasticity of Wages with Respect to Employer Size

	Country (Data)	*Sample*	*Other Controls*	*Elasticity (SE)*	*Number of Observations*	R^2
(1)	US (CPS)	Cross-section	No	0.108 (0.004)	7854	0.10
(2)	US (CPS)	Cross-section	Yes	0.064 (0.003)	7854	0.36
(3)	UK (LFS)	Cross-section	No	0.145 (0.002)	220868	0.07
(4)	UK (LFS)	Cross-section	Yes	0.074 (0.001)	220868	0.42
(5)	UK (BHPS)	Cross-section	No	0.086 (0.002)	13365	0.09
(6)	UK (BHPS)	Cross-section	Yes	0.047 (0.002)	13365	0.49
(7)	UK (BHPS)	Panel	Yes	0.013 (0.002)	13813	0.02
(8)	UK (BHPS)	Panel movers	Yes	0.035 (0.007)	1340	0.11
(9)	UK (LFS)	Dual job holders	No	0.037 (0.007)	5342	0.01

Notes.
1. The dependent variable is the log of the hourly wage. Employer size is coded as the mid-points of the relevant bands with the open-ended top category being coded as twice the lower bound. The elasticity is the coefficient on the log of the employer size variable.
2. The other controls included are marital status, children, experience, tenure (all interacted with gender), region, race, and (for the LFS) month dummies. Standard errors are reported in parentheses.

in the United Kingdom. Introducing controls reduces the estimates by 40–50% to 0.064 in the United States and 0.074 in the United Kingdom. The fifth and sixth rows estimate the elasticity using the BHPS: the elasticity here is smaller than that found in the LFS. The seventh row estimates an equation for wage growth including change in employer size as a regressor. The estimated elasticity drops to 0.013 although it remains significantly different from zero. This might suggest that controlling for unobserved worker quality makes most of the size–wage effect disappear. However, there is good reason to think that this is an understatement of the true elasticity as the employer size variable is likely to have a lot of measurement error (think of the difficulty in answering a question about the size of your workplace)[4] and this measurement error will be compounded when we estimate a model in first-differences. To get some idea of the extent of the problem, we compared worker responses to the employer size question in the BHPS to management responses to a similar question in the 1998 UK Workplace Employee Relations Survey (WERS).[5] In the BHPS only 63% of workers who did not change jobs reported their employer being in the same size class as one year ago: for WERS the (employee-weighted) figure is 88%. For the largest workplaces (those with 1000+ workers) 97% of the WERS sample reported being in the same category the previous year and the remaining 3% were in the next category (500–999 workers). In the BHPS, only 72% reported having 1000+ employees previously and 10% reported their employer previously having less than 200 employees. It is clear that there is a lot of measurement error in reported changes in employer size from employee data. This measurement error in the change in employer size is likely to be less important for workers who change jobs as the signal to noise ratio is likely to be higher. The eighth row of table 4.2 estimates a wage growth model on a sample of movers—the estimated elasticity rises to 0.037. Similar results are reported on US data sets by Brown and Medoff (1989).

[4] Measurement error is another reason why estimates that include controls may be worse than those that do not as was originally pointed out by Griliches (1977). Intuitively, the fraction of the variation in the employer size variable that is measurement error after introducing controls is likely to be higher. In the present context, measurement error should be interpreted broadly to mean any transitory shocks to employment. For example, it may be that wages are related to a long-run measure of employer size because there are pressures which limit the extent of variation in wages (see chapter 5 for a discussion of this) but that there are year-to-year variations in employment that do not get reflected in wages. This will have the same effect as measurement error on the estimated ESWE.

[5] WERS reports the actual level of employment now and a year ago, so we converted this to the size classes used in the BHPS. As WERS only reports weights which can be used to gross to the population of workplaces with 10+ employees, we also restricted this analysis to workers who report 10+ employees in the BHPS.

TABLE 4.3
Job Mobility and Employer Size

	(1)	*(2)*	*(3)*
Log employer size	−0.054 (0.008)	−0.038 (0.009)	−0.023 (0.009)
Log wage			−0.269 (0.042)
Other controls	No	Yes	Yes
Number of observations	13928	13886	13886
Pseudo-R^2	0.005	0.1	0.11

Notes.
1. The data set used is the BHPS. The sample is all of those in continuous employment between one interview and the next. The dependent variable takes the value 1 if the individual changed jobs and 0 if they did not: a probit model is estimated.
2. The other controls are sex, race, education, experience, tenure, region, marital status, children, and year dummies.

The LFS offers another way to control for individual fixed effects as, for those who have more than two jobs, it asks questions about earnings and employer size for both jobs.[6] The ninth row of table 4.2 regresses the difference in log wages on the difference in log workplace size; the estimated elasticity is 0.037, identical to that obtained from the BHPS movers. These estimates suggest that controlling for worker quality does reduce the size–wage effect but it remains significantly different from zero, implying a gap of about 10% in wages between the 75th and 25th percentile of workplace size: this is similar to the magnitudes reported by Brown and Medoff (1989).

One other hypothesis to explain the size–wage effect is that it is the result of a compensating wage differential. It may be that people dislike working in large firms per se or that other working conditions tend to be worse in large firms. Observed indicators of working conditions do not suggest that large firms tend to be worse places to work but these indicators are far from perfect. Perhaps the simplest way to test the compensating wage differentials hypothesis is to examine quit rates. If the size–wage effect is simply a compensating wage differential, utility would be equalized across firms of different sizes so there is no reason to believe that quit rates would differ by firm size. In fact, workers are much less likely to leave large employers. Table 4.3 presents some evidence from the BHPS on this point. The sample is those in continuous employment where the dependent variable takes the value 1 if the individual changed jobs and 0 if they did not. In column (1) we simply estimate a probit model with the log of employer size as a regressor. It

[6] Lemieux (1998) was the first to use this approach as a way to estimate the union wage mark-up.

is significantly negative. The second column then introduces some extra controls. The impact of employer size is reduced but still significant. Finally, the third column also introduces the log wage. It is not clear whether the wage should be included or not as the impact of employer size on quits should be through the wage. Again, the coefficient on employer size is reduced but remains significantly different from zero. This evidence suggests that workers are better off in large firms, which suggests that the size–wage correlation cannot be explained solely by compensating wage differentials.

The evidence discussed so far suggests that the size–wage effect cannot be readily explained by a competitive model of the labor market. However, this does not prove that an upward-sloping supply of labor to the individual employer is the correct explanation: there are other potential non-competitive explanations, for example, efficiency wage or rent sharing theories. While efficiency wage theories are often too vague to test, rent sharing is a tighter idea. The hypothesis is that workers manage to get a share of the rents and successful firms with large rents tend to have high levels of employment. If this is the case, we would expect that workers are better at extracting a share of the rents when they are unionized than when they are not so we would expect to see a larger employer size–wage effect in the union sector. In fact, the opposite is the case. Table 4.4 shows that in the CPS, the LFS and the BHPS the size–wage effect is 3–4 times larger in the non-union sector than the union sector (for similar findings, see Brown and Medoff 1989,

TABLE 4.4
The ESWE in Union and Non-union Sectors

	Country (Data)	*Sample*	*Other Controls*	*Elasticity (SE)*	*Number of Observations*	R^2
(1)	US (CPS)	Union cross-section	Yes	0.019 (0.008)	1231	0.30
(2)	US (CPS)	Non-union cross-section	Yes	0.067 (0.004)	6623	0.37
(3)	UK (LFS)	Union cross-section	Yes	0.017 (0.003)	22737	0.42
(4)	UK (LFS)	Non-union cross-section	Yes	0.086 (0.003)	24135	0.4
(5)	UK (BHPS)	Union cross-section	Yes	0.018 (0.002)	6619	0.47
(6)	UK (BHPS)	Non-union cross-section	Yes	0.077 (0.003)	6746	0.49

Notes.
1. As for table 4.2.

Mellow 1982).[7] This seems strong evidence against the rent sharing hypothesis; Green et al. (1996) provide a more extensive discussion of this for UK data.

This section has discussed competitive and rent sharing arguments for the employer size–wage effect. The employer size–wage effect survives all of these attempts to explain it away. The following section returns to the question of how we might try to estimate the true value of ε.

4.3 Reverse Regressions

The estimates so far suggest that, while an upward-sloping labor supply curve is a plausible explanation for part of the employer size–wage effect, the elasticity of wages with respect to employment is small (perhaps about 0.04) after we have controlled for observed and unobserved worker quality. Using the formula for the Hicks–Pigou rate of exploitation in (2.3) this elasticity is also the proportionate amount by which wages fall short of marginal product so an elasticity of 0.04 suggests only small deviations from perfect competition. But, there is no reason why one might not run a regression of log employment on the log wage (and the x variables), hope to estimate ε_{Nw} from the coefficient on wages in this regression, and then invert the elasticity to give ε. We will term this the reverse regression after the discussion in Goldberger (1984).

This regression provides an estimate of $E(\log(N_i) \mid \log(w_i), x_i)$. The following proposition tells us what we would expect to find in this case.

Proposition 4.2. *Running a regression of* $\log(N_i)$ *on* $(\log(w_i), x_i)$ *estimates*

$$E\big(\log(N_i) \mid \log(w_i), x_i\big) = \frac{1 - \rho'(\varepsilon + \eta)}{\varepsilon} \log(w_i) + \rho' \ln(1 + \varepsilon) - \frac{\beta_b - \rho'(\varepsilon\beta_a + \eta\beta_b)}{\varepsilon} x_i \tag{4.9}$$

where

$$\rho' \equiv \frac{\varepsilon\sigma_{ab} + \eta\sigma_b^2}{\varepsilon^2\sigma_a^2 + \eta^2\sigma_b^2 + 2\varepsilon\eta\sigma_{ab}} \tag{4.10}$$

Proof. See Appendix 4A.

[7] It is also worth noting that Teulings and Hartog (1998) find that the ESWE is smaller in "corporatist" countries where the level of wage bargaining is above that of the individual employer.

In the reverse regression, OLS only gives unbiased estimates of $(1/\varepsilon)$ if $\varepsilon\sigma_{ab} = -\eta\sigma_b^2$. But, of course, the bias will generally be different from that in the wage equation; compare (4.9) and (4.7). If $\sigma_{ab} = 0$, then the true estimate of ε must lie between that estimated by a regression of $\log(w)$ on $\log(N)$, and one obtained from a regression of $\log(N)$ on $\log(w)$.

Now compare the direct and reverse regressions. Table 4.5 presents some further estimates of (4.7) and (4.9) using data from the US CPS and the UK LFS that have already been used in table 4.2. In each row we report the coefficient on employer size when (4.7) is estimated in the column headed "coefficient on log employer size" while the inverse of the coefficient on the log wage when (4.9) is estimated is reported in the column headed "inverse of coefficient on log wage." As these are both estimates of ε, one would hope that these coefficients are similar.

In fact, they are very different. Consider the US results first. The first row reports the results when no other controls are included and subsequent rows add personal controls, education controls, regional controls, industry controls, and occupation controls, finishing with a model with all variables included. There are several general conclusions. First, for both the United States and the United Kingdom, there is always a large gap between the coefficient on log employer size and the inverse of the coefficient on the log wage. For example, in the first row of table 4.5, the second column suggests an elasticity of the wage with respect to employment of 0.116 while the third suggests an elasticity of 1.000. Secondly, the inclusion of controls always reduces the coefficient in the second column and raises the coefficient in the third column.

What is the explanation for this pattern of results? The gap in the two estimates of ε suggests the presence of the biases identified in (4.7) and (4.9) or of measurement errors in wages or employment, or both. But, why should this gap widen when controls are introduced?

As discussed earlier, the biases in (4.7) and (4.9) depend on relative variances and the correlation coefficient of the errors in the labor supply and MRPL equations, and in a complicated way. It is possible that the biases move in different directions when the introduction of controls changes the relative variances and the correlation coefficient but it is hard for intuition to deliver any firm expectation.

The measurement error argument of Griliches (1977) is perhaps more plausible so some consideration of the likely sources of measurement error in both employer size and wages is likely to be worthwhile. For employer size, individuals have no reason to know the size of their workplace and seem to often make big mistakes as the earlier discussion has made clear. If this is the case, one would expect the coefficient on a better measure of employer size in the direct regression to be larger. It is plausible to believe that managers in workplaces have better information than

TABLE 4.5
Direct and Reverse Regressions

Sample	*Coefficient on Log Employer Size*	*Inverse of Coefficient on Log Wage*	*Personal Controls*	*Education Controls*	*Regional Controls*	*Industry Controls*	*Occupation Controls*
US	0.116 (0.004)	1.000 (0.033)	No	No	No	No	No
US	0.099 (0.004)	1.070 (0.041)	Yes	No	No	No	No
US	0.071 (0.004)	1.273 (0.065)	Yes	Yes	No	No	No
US	0.095 (0.004)	1.110 (0.044)	Yes	No	Yes	No	No
US	0.085 (0.004)	1.392 (0.067)	Yes	No	No	Yes	No
US	0.078 (0.004)	1.181 (0.055)	Yes	No	No	No	Yes
US	0.060 (0.004)	1.608 (0.101)	Yes	Yes	Yes	Yes	Yes
UK	0.144 (0.003)	1.908 (0.037)	No	No	No	No	No
UK	0.115 (0.003)	1.988 (0.045)	Yes	No	No	No	No
UK	0.092 (0.002)	2.070 (0.054)	Yes	Yes	No	No	No
UK	0.110 (0.002)	1.964 (0.045)	Yes	No	Yes	No	No
UK	0.085 (0.003)	2.671 (0.088)	Yes	No	No	Yes	No
UK	0.075 (0.002)	2.368 (0.073)	Yes	No	No	No	Yes
UK	0.062 (0.002)	2.737 (0.109)	Yes	Yes	Yes	Yes	Yes

Notes.
1. US data are from the Contingent Worker Survey to the April 1993 CPS. The sample is restricted to the non-union sector. UK data are from the LFS.
2. The dependent variable is the log of the hourly wage.
3. Personal controls are sex, race, a quartic in experience, marital status, and the presence of children.
4. The column headed "coefficient on log employer size" gives the results from estimating a regression of log wages on log employer size and other controls. The column headed "inverse of coefficient on log wage" gives the results from estimating a regression of log employer size on log wages and other controls.

TABLE 4.6
The ESWE: Evidence from WERS

	Coefficient on Log Employer Size	*Inverse of Coefficient on Log Wage*	*Controls*	*Method of Estimation (Instrument)*
(1)	0.120 (0.010)	–	No	OLS
(2)	0.051 (0.007)	–	Yes	OLS
(3)	0.050 (0.007)	–	Yes	IV (first lag)
(4)	0.056 (0.007)	–	Yes	IV (fifth lag)
(5)	–	3.383 (0.365)	Yes	OLS
(6)	–	1.408 (0.188)	Yes	Between-firm

Notes.

1. Data are from the 1998 UK WERS. The number of observations is approximately 25,000. The dependent variable is the log of the hourly wage for rows (1)–(4) and the log of employers size in rows (5) and (6).
2. The controls included are sex, race, education, age, occupation, and industry. Observations are weighted to be representative of all workers in plants with more than 10 employees (smaller plants are excluded). Standard errors are computed assuming clustering on the workplace.

their workers about employer size, although it might be better to have administrative data like that used by Bayard and Troske (1999). The 1998 UK WERS can be used to investigate this. Some results are reported in table 4.6. The first row reports the result of a simple regression of log wages on log employer size. Log wages are reported by the individual workers concerned and employer size by the manager so these regressions are very similar to those reported earlier except that employer size is manager-reported. However, the coefficient on log employer size is similar to what we have seen before. The second row introduces controls: the drop in the coefficient on log employer size is very similar to what we have seen before. There is little evidence here to suggest substantial worker misreporting of employer size although the earlier results on the excessive apparent changes in employer size in the BHPS did suggest the existence of large measurement errors.[8]

An additional reason for why the coefficient on log employer size may be underestimated is the existence of transitory shocks to employment that are not reflected in wages. WERS also allows us a way to investigate this as employers are asked to report workplace employment one and five years previously. A simple regression of the log of employment this year on the log of employment last year has a coefficient of 0.97 on the lagged dependent variable suggesting a lot of permanence in the level of employment. The

[8] It is possible that the comparison of BHPS and WERS results is made difficult by the different nature of the two data sets.

third row of table 4.6 shows the coefficient on $\log(N)$ when it is instrumented by lagged employment to try to pick up the permanent component in employment. The instrumental variable coefficient is very similar to the ordinary least squares. The fourth row uses the fifth lag of employment as an instrument with very similar results. These results suggest that transitory shocks to employment that are not reflected in wages are not particularly important in explaining the estimated size of the ESWE.

The discussion so far has focused on measurement errors and transitory shocks to employment. But, there are reasons to think there might be similar problems surrounding wages. Perhaps the most important is that the wage used in the regressions in table 4.2 refers to the wage of a single worker in the plant. This is an unbiased estimate of the average log wage in the plant but obviously contains some measurement error. As a result, the coefficient on log wages in the reverse regression is likely to be biased towards zero. Evidence for this effect can be seen in the fifth and sixth rows of table 4.6. The fifth row reports the result of a reverse regression on the WERS data. The implied value of ε is very high. The sixth row reports the result of a between-plant regression on the same data, exploiting the fact that the data set contains observations on multiple workers within plants. This effectively runs a regression of the log of employment on the average log wage. The coefficient on the wage variable rises implying a lower value of ε, as one would expect.

None of the discussion so far gives us much confidence that OLS, either a direct or reverse regression, gives us a good estimate of ε. On the basis of the results reported, one could argue either that labor supply to the firm is very inelastic or that it is quite elastic. It is perhaps better to conclude that these regressions are just not very informative. But, while it might have been nice for OLS to deliver reliable results, perhaps it was expecting too much and a simpler potential solution presents itself.

To identify the labor supply curve (which is all we want here), a variable that shifts the MRPL curve without shifting the supply curve is needed. One can then use this as an instrument for the wage or employment in estimating the supply curve (depending on which way round we are estimating the supply curve). This procedure will yield consistent estimates of ε. But, of course, it requires us to be able to provide such an instrument.

If one is interested in estimating the elasticity of labor supply to an individual firm, then the instrument needs to be something that affects the demand curve for that firm but has negligible impact on the labor market as a whole. The reason is that a pervasive labor market demand shock will raise the general level of wages so is likely to be correlated with B in (4.2). So, for example, the approach of using demand shocks caused by exchange rate fluctuations (as in Abowd and Lemieux 1993) does not seem viable here.

There are a number of studies that attempt to use firm-level instruments. For example, Sullivan (1989) uses the population in the area surrounding the hospital as an instrument affecting the demand for nurses, and Beck et al. (1998) use the number of children in a school district as an instrument for the demand for teachers. These represent serious attempts to deal with a difficult problem but their instruments are not beyond criticism. If the main variation in the number of children or the number of patients comes from variation in population, it is also likely that the supply of nurses and teachers in an area is proportional to population as well. Perhaps the best studies are Staiger et al. (1999) and Falch (2001). Staiger et al. (1999) examine the impact of a legislated rise in the wages paid at Veteran Affairs hospitals. This combined with a plausible argument that these hospitals were allowed to hire as many staff as they wanted (which is required to make sure we are estimating the supply curve) seems as close to an exogenous increase in wages as anything else in the literature. They estimate the short-run elasticity in the labor supply to the firm to be very low—around 0.1 implying an enormous amount of monopsony power possessed by hospitals over their nurses. Falch (2001) investigates the impact on the supply of teachers to individual schools in Norway in response to a policy experiment that selectively raised wages in some schools. Again, he finds that the elasticity in the supply of labor to individual schools is very low. How plausible are these estimates and whether they can be generalized to the rest of the labor market are open questions as the markets for nurses and teachers are ones that might conventionally be thought of as having some monopsonistic elements.

This is all rather depressing: a good estimate of the elasticity of the labor supply curve facing the firm seems very elusive so perhaps there is a very good reason for the lack of research into this area. Progress seems to be dependent on finding a good firm-level instrument.

4.4 Estimating Models of Dynamic Monopsony

The previous part of this chapter has used the static model of monopsony as a way to think about the issue of estimating the elasticity of the labor supply curve facing an individual firm. The rest of this chapter uses a more explicitly dynamic, theoretical approach to estimate this elasticity. In a steady state, we know that the supply of labor to the firm $N(w)$ must be given by $N(w) = R(w)/s(w)$ where $R(w)$ is the flow of recruits to the firm and $s(w)$ is the separation rate. As pointed out by Card and Krueger (1995) and discussed earlier in section 2.2, this implies that

$$\varepsilon_{Nw} = \varepsilon_{Rw} - \varepsilon_{sw} \tag{4.11}$$

so that knowledge of the elasticities of recruitment and quits with respect to the wage can be used to estimate the elasticity of labor supply facing the firm. This section discusses how we can estimate ε_{Rw} and ε_{sw}.

One of the advantages of this approach is that there is a well-established literature that discusses the elasticity of the separation rate with respect to the wage (e.g., Pencavel 1972; Parsons 1972, 1973; Viscusi 1980; Light and Ureta 1992). However, an apparent disadvantage is that it might be unclear how the elasticity of the recruits with respect to the wage should be estimated. Card and Krueger (1995) use estimates of the elasticity of job applicants with respect to the wage from Holzer et al. (1991) and Krueger (1988) but the justification for this is not obvious. One of the contributions here is to show that there is a close connection between ε_{Rw} and ε_{sw} so that estimates of the separation elasticity are informative about the recruitment elasticity.

Consider the basic Burdett and Mortensen (1998) model of dynamic monopsony introduced in section 2.4. In this model, we have

$$s(w) = \delta + \lambda[1 - F(w)] \tag{4.12}$$

$$R(w) = R^{u} + \lambda \int^{w} f(x)N(x)dx \tag{4.13}$$

where $s(w)$ is the separation rate in a firm that pays wage w, $R(w)$ is the flow of recruits, δ is the rate (assumed exogenous) at which workers leave employment for non-employment, λ is the arrival rate of job offers, $F(w)$ is the distribution of wage offers, R^{u} are the recruits from unemployment (which does not depend on the wage offered) and $N(w)$ is the employment level in a firm that pays w. By differentiating (4.12) and (4.13), we have

$$\varepsilon_{sw} = \frac{ws'(w)}{s(w)} = -\frac{\lambda wf(w)}{s(w)} = -\frac{\lambda wf(w)N(w)}{R(w)} = -\frac{wR'(w)}{R(w)} = -\varepsilon_{Rw} \tag{4.14}$$

where the third equality sign follows from the fact that, in steady state, $s(w)N(w) = R(w)$. (4.14) says that, in a steady state, the recruitment elasticity is simply minus the separation elasticity so that (using (4.11)) one can simply double the separation elasticity to get an estimate of the labor supply elasticity. The explanation for the connection between the two elasticities is that separations from one firm for a wage-related reason must be the recruit of some other firm so that quits and recruits are two sides of the same coin.

Although this result is neat, one might wonder about its robustness. So, let us consider some generalizations. First, relax the assumption that workers always quit for a better-paying job and never quit for a job with lower pay. Suppose that a worker currently being paid w accepts a job offer of x with probability $\phi(x/w)$. We assume that it is the ratio of the wages that matters which seems a reasonable restriction as there is no reason to think that a general increase in wages would have any effect on job-to-job mobility rates. Note that the model of (4.12) and (4.13) corresponds to the case where $\phi(x/w) = 0$ if $x < w$ and $\phi(x/w) = 1$ if $x > w$. The separation rate and recruitment functions will now be given by

$$s(w) = \delta + \lambda \int \phi\left(\frac{x}{w}\right) f(x)dx \tag{4.15}$$

$$R(w) = R^{\mathrm{u}} + \lambda \int \phi\left(\frac{w}{x}\right) f(x)N(x)dx \tag{4.16}$$

The following proposition tells us that a suitably weighted separation elasticity must be equal to a suitably weighted recruitment elasticity.

Proposition 4.3. *If the separation and recruitment functions are given by (4.15) and (4.16), then the recruit-weighted separation and recruitment elasticities must be equal, that is,*

$$\int \varepsilon_{sw}(w)R(w)f(w)dw = -\int \varepsilon_{Rw}(w)R(w)f(w)dw \tag{4.17}$$

Proof. See Appendix 4A.

If the separation and recruitment elasticities are both constant, (4.17) says that they must be equal. If they vary with the wage then they can differ but probably not by much. For example, if the separation elasticity is finite everywhere, the recruitment elasticity cannot be infinite for any positive measure of employees.

However, the separation and recruitment functions of (4.15) and (4.16) are still quite restrictive in that they assume that separations to and recruitment from non-employment are not sensitive to the wage.

Write total separations as $s(w) = s^{\mathrm{n}}(w) + s^{\mathrm{e}}(w)$ where $s^{\mathrm{n}}(w)$ is the separation rate to non-employment and $s^{\mathrm{e}}(w)$ is the separation rate to employment. Denote by θ_s the share of separations which are a direct move to another job. Similarly write total recruits as $R(w) = R^{\mathrm{n}}(w) + R^{\mathrm{e}}(w)$ where $R^{\mathrm{n}}(w)$ is the flow of recruits from non-employment and $R^{\mathrm{e}}(w)$ is the flow of recruits from employment. Denote by θ_R the share of recruits from employment. The overall elasticity of labor supply with respect to the wage can be written as

$$\varepsilon_{Nw} = \theta_R \varepsilon^{\mathrm{e}}_{Rw} + (1 - \theta_R)\varepsilon^{\mathrm{n}}_{Rw} - \theta_s \varepsilon^{\mathrm{e}}_{sw} - (1 - \theta_s)\varepsilon^{\mathrm{n}}_{sw} \tag{4.18}$$

so that knowledge of the four elasticities can be used to compute the elasticity of the labor supply curve facing the firm. The two separation elasticities can be estimated straightforwardly (see below) so the only problem is how to estimate the recruitment elasticities. The following proposition shows that a suitably weighted recruitment elasticity from employment is equal to a weighted separation elasticity to employment.

Proposition 4.4. *If the separation and recruitment functions to and from employment are given by*

$$s^{\mathrm{e}}(w) = \lambda \int \phi\left(\frac{x}{w}\right) f(x)dx \tag{4.19}$$

$$R^{\mathrm{e}}(w) = \lambda \int \phi\left(\frac{w}{x}\right) f(x)N(x)dx \tag{4.20}$$

then a suitably weighted average of the separations elasticity must be equal to a weighted average of the recruitment elasticities. In particular,

$$\frac{\int \varepsilon^{\mathrm{e}}_{sw}(w)s^{\mathrm{e}}(w)N(w)f(w)dw}{\int s^{\mathrm{e}}(w)N(w)f(w)dw} = -\frac{\int \varepsilon^{\mathrm{e}}_{Rw}(w)R^{\mathrm{e}}(w)f(w)dw}{\int R^{\mathrm{e}}(w)f(w)dw} \tag{4.21}$$

Proof. See Appendix 4A.

(4.21) says that a weighted average of the separation elasticity and the recruitment elasticity must be equal but that the weights are not equal, as in (4.17).[9] However, this still means that if both the separation and recruitment elasticities are constant, they must be equal.

Unfortunately, there is no equivalent result relating the separation and recruitment elasticities for transitions to and from non-employment. For example, if there is heterogeneity in the reservation wages of workers but the reservation wage for an individual worker never changes, recruits from non-employment are increasing in the wage but separations to non-employment are not. However, if there is a stochastic component to the reservation wage, separations to non-employment are also sensitive to the wage. The basic problem is that, whereas a move from one job to another is a quit for one firm and immediately a recruit for another firm, this is not true of flows between employment and non-employment.

So, we cannot use the wage elasticity of separations to non-employment to estimate the wage elasticity of recruits from non-employment: we need a different method. The share of recruits from employment is given by

[9] The separation elasticity is weighted by separations to other jobs while the recruitment elasticity is weighted by recruits from other jobs. We would expect the weight on the elasticity in high-wage firms to be larger for the recruitment than the separation elasticity.

$$\theta_R(w) = \frac{R^{\mathrm{e}}(w)}{R^{\mathrm{e}}(w) + R^{\mathrm{n}}(w)} \tag{4.22}$$

This enables us to prove the following relationship between $\varepsilon_R^{\mathrm{e}}(w)$ and $\varepsilon_R^{\mathrm{n}}(w)$.

Proposition 4.5. *If the share of recruits from employment is given by (4.22), then the relationship between the wage elasticity of recruits from non-employment,* $\varepsilon_{Rw}^{\mathrm{n}}(w)$*, and the wage elasticity of recruits from employment,* $\varepsilon_{Rw}^{\mathrm{e}}(w)$*, is given by*

$$\varepsilon_{Rw}^{\mathrm{n}}(w) = \varepsilon_{Rw}^{\mathrm{e}}(w) - \frac{w\theta_R'(w)}{\theta_R(w)[1 - \theta_R(w)]} \tag{4.23}$$

Proof. See Appendix 4A.

If, for example, we model $\theta_R(w)$ as a logistic function $e^{\beta x}/(1 + e^{\beta x})$ where x includes the log wage, then $(w\theta_R'/\theta_R(1 - \theta_R)) = \beta_w$ where β_w is the coefficient on the log wage.

Summarizing all this information, our strategy for estimating the elasticity of the labor supply curve facing the firm is

- estimate separations equations for separations to employment and non-employment, and obtain the wage elasticities;
- use the wage elasticity of separations to employment to estimate the wage elasticity of recruits from employment (based on Proposition 4.4);
- estimate a logit model for the probability that a recruit comes from employment and then use (4.23) to estimate the elasticity of recruits from non-employment;
- use these elasticities and information of the share of separations to and recruits from employment in (4.14) to estimate the elasticity of the labor supply curve facing the firm.

Let us now put this into practice.

4.5 Estimating the Wage Elasticity of Separations

We model the instantaneous separation rate as $s = e^{\beta x}$. One of the x variables will be the log of the wage so that the elasticity of the separation rate with respect to the wage will simply be the coefficient on the wage. From the previous discussion, we also need to model the separations to employment and non-employment separately. Write the separation rate to other jobs as $s^{\mathrm{ee}}(x) = \exp(\beta^{\mathrm{ee}}x)$ and the separation rate to non-employment as $s^{en}(x) = \exp(\beta^{en}x)$. We assume that, conditional on x, the two sorts of separation are independent.

Define an indicator variable y^{en} which takes the value 1 if the individual has a spell of non-employment in a period of time τ and 0 otherwise and an another indicator variable y^{ee} which, if the individual does not have a spell of non-employment, takes the value 1 if the individual changes jobs and 0 if they do not. The probabilities of the different outcomes are given by

$$\Pr(y^{en} = 1 \mid x) = 1 - \exp(-s^{en}(x)\tau)$$

$$\Pr(y^{en} = 0, y^{ee} = 1 \mid x) = \exp(-s^{en}(x)\tau)(1 - \exp(-s^{ee}(x)\tau)) \qquad (4.24)$$

$$\Pr(y^{en} = 0, y^{ee} = 0 \mid x) = \exp(-s^{en}(x)\tau)\exp(-s^{ee}(x)\tau)$$

so that the individual contribution to the log-likelihood function can be written as

$$\log L = y^{en}\ln[1 - \exp(-s^{en}(x)\tau)] + (1 - y^{en})\ln[\exp(-s^{en}(x)\tau)]$$

$$+(1 - y^{en})[y^{ee}\ln[1 - \exp(-s^{ee}(x)\tau)] + (1 - y^{ee})\ln[\exp(-s^{ee}(x)\tau)]] \qquad (4.25)$$

The important feature of (4.25) is that one can estimate the separations elasticity to non-employment and other jobs separately. To estimate the elasticity of separations to non-employment, the whole sample is used and we have as a dependent variable whether the individual had a period of non-employment in the year. To estimate the elasticity of separations to other jobs the sample of those who have been in continuous employment is used and we have as a dependent variable whether the individual remains in the same job. Note that the overall elasticity will be a weighted average of these two elasticities, the weight being the fraction of separations that are to non-employment.

The most serious problem in estimating the wage elasticities is, as always, going to be the result of a failure to control adequately for other relevant factors. One potential source of problems in estimating the separation elasticity is a failure to control adequately for the average level of wages in the individual's labor market. Separations are likely to depend on the wage relative to this alternative wage so that a failure to control for the alternative wage is likely to lead to a downward bias in the wage elasticities. On the other hand, we would expect separations to be more sensitive to the permanent component of wages than to the part of wages that is a transitory shock or measurement error. In this case, the inclusion of controls correlated with the permanent wage is likely to reduce the estimated wage elasticity. Table 4.7 estimates some separations equations with and without controls for the PSID, NLSY, BHPS and

TABLE 4.7
The Sensitivity of the Separation Elasticity to Specification

	PSID (US)	*NLSY (US)*	*BHPS (UK)*	*LFS (UK)*
All separations				
Mean separation rate	0.21	0.55	0.19	0.058
No controls	−0.944 (0.030)	−0.515 (0.019)	−0.798 (0.032)	−0.646 (0.021)
With controls	−0.973 (0.041)	−0.536 (0.032)	−0.720 (0.041)	−0.500 (0.028)
Tenure controls	−0.575 (0.037)	−0.340 (0.026)	−0.503 (0.064)	−0.343 (0.032)
Separations to employment				
Mean separation rate	0.12	0.43	0.12	0.032
No controls	−0.759 (0.050)	−0.307 (0.018)	−0.631 (0.038)	−0.529 (0.030)
With controls	−0.867 (0.038)	−0.359 (0.032)	−0.688 (0.049)	−0.425 (0.039)
Tenure controls	−0.450 (0.042)	−0.156 (0.027)	−0.429 (0.050)	−0.207 (0.044)
Separations to non-employment				
Mean separation rate	0.08	0.12	0.07	0.025
No controls	−1.010 (0.067)	−0.750 (0.028)	−0.916 (0.048)	−0.748 (0.029)
With controls	−0.892 (0.087)	−0.850 (0.055)	−0.632 (0.066)	−0.578 (0.041)
Tenure controls	−0.569 (0.068)	−0.713 (0.056)	−0.493 (0.071)	−0.477 (0.045)

Notes.
1. This table reports the elasticities of separations with respect to the wage. The PSID, NLSY, and BHPS samples are those described in the Data Sets Appendix. The LFS sample is from September 1997 to November 1999. The LFS also differs from the other data sets in modeling labor market transitions from one quarter to another instead of one year to another. This is why the means of the dependent variables are so much lower. The row headed "no controls" simply includes the wage. The rows marked "with controls" include gender, education, race, marital status, children, region, a quartic in experience, and year dummies. The row headed "tenure controls" includes a quartic in tenure in addition to the usual controls.

LFS. First consider the wage elasticity for all separations (the top panel of table 4.7). All the estimated wage elasticities are negative and significantly different from zero. They range from −0.5 for the NLSY to −0.9 for the PSID. The bottom two panels estimate separate wage elasticities for separations to employment and non-employment. There is evidence that the elasticities for both separations to employment and non-employment are both sensitive to the wage with the latter being larger than the former.

Inclusion of standard human capital controls does not make much difference to the estimated wage elasticities. However, one variable whose inclusion or exclusion makes a lot of difference to the apparent estimated wage elasticity is job tenure.[10] The inclusion of job tenure always drastically reduces the estimated wage elasticity as high-tenure workers are less likely to leave the firm and are more likely to have high wages. There are arguments both for and against the inclusion of job tenure. One of the benefits of paying high wages is that tenure will be higher so that one needs to take account of this indirect effect if one wants the overall wage elasticity when including tenure controls: in this situation, excluding tenure may give better estimates. On the other hand, if there are seniority wage scales, the apparent relationship between separations and wages may be spurious.

Unobserved heterogeneity that is correlated with the wage causes familiar problems but the wage elasticity is likely to be biased even if there is heterogeneity uncorrelated with the wage. To see this, suppose that the separation rate is $\xi w^{-\beta}$ where ξ is unobserved and independent of the wage. To keep things tractable we will assume that ξ has a gamma distribution with mean μ and variance σ^2. The following proposition summarizes the effect of unobserved heterogeneity on the estimated wage elasticity.

Proposition 4.6. *The elasticity of the separations rate with respect to the wage is biased towards zero with gamma-distributed unobserved heterogeneity that is uncorrelated with the wage. The shorter the time period over which the data are observed, the smaller is the bias.*

Proof. See Appendix 4A.

The result on the existence of a bias is unsurprising: we are using survivor functions to estimate the wage elasticity and it is well known that unobserved heterogeneity has an effect on the estimated coefficients in duration models (see Lancaster 1990). Table 4.8 investigates whether the time horizon makes any difference to the estimated wage elasticity

[10] The word "apparent" is appropriate here because the dependence of job tenure on the wage needs to be taken into account when estimating the full wage elasticity.

TABLE 4.8
The Effect of the Time Horizon on the Separation Elasticity: UK LFS

Controls	*No*	*Yes*
1 quarter	−0.646 (0.021)	−0.500 (0.028)
2 quarters	−0.640 (0.018)	−0.497 (0.024)
3 quarters	−0.586 (0.017)	−0.471 (0.023)
4 quarters	−0.547 (0.017)	−0.429 (0.023)

Notes.
1. The reported coefficients are the coefficients on the log wage from the estimated separations model as described in section 4.5. The controls included are gender, race, education, experience, martial status and dependent children, region, and month.
2. The dependent variable in the row marked 1 quarter is whether the individual left the initial job over the first quarter, that for 2 quarters whether the individual left over the first two quarters, etc.

using data from the UK LFS. Results are reported for the wage elasticity estimated over a period of one to four quarters. The results are consistent with the predictions of Proposition 4.6. Both with and without other controls, the estimated wage elasticity is higher over short periods than long. As all the other data sets used in table 4.3 use a time horizon of a year, this suggests the estimates may be understating the true wage elasticity. However, the results in table 4.8 do not suggest the size of the bias is large. The estimated elasticity of separations with respect to the wage rises (in absolute terms) from −0.43 to −0.5 as the time horizon is narrowed from one year to one quarter and hardly narrows at all as the time horizon falls from two quarters to one quarter.

4.6 The Proportion of Recruits from Employment

Proposition 4.5 says that we need to know how the fraction of recruits from employment varies with the wage. Table 4.9 reports the results of estimating a logit model for a recruit coming from employment for our four data sets. In all data sets, the higher the wage, the higher the probability that a recruit comes from employment: this is as the theory predicts. This implies that the wage elasticity of recruits from employment is higher than the wage elasticity of recruits from non-employment.

4.7 The Elasticity of the Labor Supply Curve Facing the Firm

We are now in a position to provide an estimate of the labor supply curve facing the firm using the results of the previous two sections and (4.18).

TABLE 4.9
The Probability of a Recruit Coming from Employment

Data Set	*Mean of Dependent Variable*	*Coefficient (SE) on Log Wage Without Controls*	*Coefficient (SE) on Log Wage with Controls*	*Number of Observations*
PSID	0.29	1.011 (0.036)	0.948 (0.054)	14277
NLSY	0.32	0.533 (0.037)	0.674 (0.042)	13653
BHPS	0.36	1.129 (0.065)	1.384 (0.080)	4649
LFS	0.51	0.824 (0.035)	0.746 (0.042)	12071

Notes.
1. The dependent variable is a dummy variable taking the value 1 if a worker was recruited from employment and 0 otherwise. The sample is all recruits. The estimation method is logit. The other controls are gender, race, experience, education, region, and year dummies.

The results of this are reported in table 4.10 where no attempt has been made to correct the wage elasticities for the problems caused by measurement error and the time horizon. These labor supply elasticities are low—in the region of 1. There are a number of reasons why one might argue that these elasticities are underestimates but it is clear that extremely large adjustments to these estimates are necessary to make perfect competition an acceptable approximation. For example, the evidence on the size of the bias caused by the interaction of the time horizon and unobserved heterogeneity would not dramatically increase these elasticities. One can only conclude that the elasticity of separations with respect to the wage is low and that this results in a low elasticity in the supply of labor to individual employers.

TABLE 4.10
The Elasticity of the Labor Supply Curve Facing the Firm

Data	*PSID*	*NLSY*	*BHPS*	*LFS*
Elasticity of separations to employment	0.867	0.359	0.631	0.529
Elasticity of separations to non-employment	0.892	0.850	0.632	0.578
Share of separations to employment	0.62	0.78	0.63	0.56
β_w	0.948	0.674	1.384	0.746
Elasticity of labor supply curve	1.38	0.68	0.75	0.75

Notes.
1. These computations used table 4.7 for the separation elasticities and the share of separations to employment, table 4.9 for the estimate of β_w, and (4.18) for the elasticity of the labor supply curve. The share of recruits from employment is assumed to be equal to the share of separations to employment as must be the case in steady state.

4.8 The Estimation of Structural Equilibrium Search Models of the Labor Market

This is the best place in this book to discuss a small but relevant literature on the structural estimation of equilibrium search models. The earliest empirical model, Eckstein and Wolpin (1990), used the Albrecht and Axell (1984) model but most later contributions have based their empirical analysis on some variant of the Burdett and Mortensen (1998) model described in section 2.4 (for surveys, see van den Berg 1999; Mortensen 2002). Because the parameters of that model contain all the necessary information (assuming, of course, that the model is correctly specified) for working out the supply of labor facing the firm, these estimates contain within them an estimate of the wage elasticity of the labor supply to an individual firm. Sometimes this elasticity is made explicit, although more often it is not.

Whether this general equilibrium approach to empirical modeling is a superior methodology depends on the purpose to which one is going to put the estimates. If one wants to model the general equilibrium effects of a change in an economy-wide policy, then such an approach may be essential. But if one simply wants an estimate of the extent of employer market power there are reasons to think that the structural approach offers few advantages and many disadvantages. The earliest estimates of equilibrium search models (e.g., Kiefer and Neumann 1993; van den Berg and Ridder 1998) emphasized how the general equilibrium model provided a tight link between the wage distribution and labor market transition rates that was exploited in the structural estimation. But, this link was more of a problem than a help and the models did not explain the distribution of wages well. More sophisticated models were introduced that added employer and worker heterogeneity (see Bontemps et al. 1999, 2000) and generalizing the wage policy used by employers (see Postel-Vinay and Robin 2002). The effect of these reasonable generalizations is effectively to allow the distribution of wages to vary independently of the labor market frictions. In this case, joint estimation of transition rates and the wage distribution offers no real advantages, and the complication of the models hides the simple economics at work. For all their sophistication one ends up with estimates of the extent of frictions that are not much more advanced than the back-of-the-envelope calculations in section 2.6. And, the partial equilibrium approach described in this chapter dominates that approach as it makes less in the way of stringent assumptions about the economy. So, it is probably the case that these equilibrium search models have told us little about the extent of frictions in the labor market that we could not have learned in their absence.

The discussion here mirrors a wider debate in econometrics about the merits of structural modeling. Structural models provide excellent estimates if the model is correct but may not be robust to small deviations from the maintained model, and tractability may restrict attention to implausibly simplistic models. A more pragmatic approach is likely to provide more robust estimates that may not be fully efficient if one pretends to know the true model. As Wolpin (1992: 558) put it 10 years ago "methods for estimating dynamic stochastic models of this kind are still in a relatively undeveloped stage, and knowledge about the effects of model and solution misspecification is very limited. ... exactly, how seriously one should take these particular estimates as reflecting real phenomena is open to debate." Unfortunately, 10 years on, the debate remains just as open and little progress has been made. For example, a state-of-the-art paper in the area (Postel-Vinay and Robin 2002) apparently shows that a model based on the totally implausible assumption of universal offer-matching (with the implication, among others, that unemployed workers are indifferent about getting a job or not) to maximize profits (in an economy, France, where union coverage approaches 100%) is consistent with the observed data, it is time to start worrying about identification as at least one other model of the labor market (the correct one) must also be consistent with the data. My gut feeling is that, for the purpose of estimating the wage elasticity in the supply of labor to an individual employer, a pragmatic approach is more likely to deliver credible results.

4.9 Conclusions

The fundamental difference between monopsony or oligopsony and perfect competition is the size of the elasticity of the labor supply curve facing a firm. Perfect competition assumes it is infinite, imperfect competition that it is finite. There is remarkably little literature on estimating the elasticity of this labor supply curve and this chapter has tried to fill that gap. It has investigated two main methods: one based on the correlation between wages and employment (the employer size–wage effect) and the other based on the estimation of separations functions. The two approaches give rather different results. OLS estimates of the ESWE suggest that the wage elasticity of the labor supply curve to individual employers is in the region of 10–15 leading to a wage that is 6–10% below marginal revenue product. However, there are reasons to think that OLS overstates the wage elasticity and IV estimates are a lot lower. The approach based on estimating the wage elasticity of separations suggests a wage elasticity of the labor supply curve to individual employers that is around one. While there are reasons to think these may be underestimates, it would certainly be hard to argue on the basis of these estimates that the labor supply elasticity is anywhere near 10. However, neither approach is entirely satisfactory: progress really needs good firm-level instruments although these are likely to be hard to find. Given this, it would probably be unwise to base one's belief in the market power of employers too much on these estimates if there were no other evidence. But, as the rest of the book sets out to show, there are very good reasons for believing that employers do have non-negligible market power.

Appendix 4A

Proof of Proposition 4.1

Taking logs of (4.2), we have

$$\begin{aligned} E\big(\log(w_i) \mid \log(N_i), x_i\big) &= E\big(b_i \mid \log(N_i), x_i\big) + \varepsilon \log(N_i) \\ &= \beta_b x_i + E\big(\nu_{bi} \mid \log(N_i), x_i\big) + \varepsilon \log(N_i) \end{aligned} \tag{4.26}$$

Now, from (4.4) and (4.6) we can derive

$$\begin{aligned} &E\big(\nu_{bi} \mid \log(N_i), x_i\big) \\ &\quad = E\big(\nu_{bi} \mid \nu_{ai} - \nu_{bi} = (\varepsilon + \eta)\log(N_i) - (\beta_a - \beta_b)x_i + \ln(1+\varepsilon)\big) \\ &\quad = \frac{\sigma_{ab} - \sigma_b^2}{\sigma_a^2 + \sigma_b^2 - 2\sigma_{ab}}\big[(\varepsilon + \eta)\log(N_i) - (\beta_a - \beta_b)x_i + \ln(1+\varepsilon)\big] \end{aligned} \tag{4.27}$$

where the second equality follows from standard results on bivariate normal distributions. This can be written as (4.7) and (4.8).

Proof of Proposition 4.2

Taking the log of (4.2), leads to

$$\begin{aligned} E\big(\log(N_i) \mid \log(w_i), x_i\big) &= \frac{1}{\varepsilon}\log(w_i) - \frac{1}{\varepsilon}E\big(b_i \mid \log(w_i), x_i\big) \\ &= \frac{1}{\varepsilon}\log(w_i) - \frac{\beta_b}{\varepsilon}x_i - \frac{1}{\varepsilon}E\big(\nu_{bi} \mid \log(w_i), x_i\big) \end{aligned} \tag{4.28}$$

Now, from (4.5) and (4.6) and standard results on the bivariate normal distribution, we can derive

$$
\begin{aligned}
&E\big(\nu_{bi} \mid \log(w_i), x_i\big) \\
&\quad = E\big(\nu_{bi} \mid \varepsilon\nu_{ai} + \eta\nu_{bi} = (\varepsilon + \eta)\log(w_i) - (\varepsilon\beta_a + \eta\beta_b)x_i + \varepsilon\ln(1+\varepsilon)\big) \\
&\quad = \frac{\varepsilon\sigma_{ab} + \eta\sigma_b^2}{\varepsilon^2\sigma_a^2 + \eta^2\sigma_b^2 + 2\varepsilon\eta\sigma_{ab}}\big[(\varepsilon + \eta)\log(w_i) - (\varepsilon\beta_a + \eta\beta_b)x_i + \varepsilon\ln(1+\varepsilon)\big]
\end{aligned}
\tag{4.29}
$$

which, substituting into (4.28) leads to (4.9) and (4.10).

Proof of Proposition 4.3

Differentiating (4.15), we have

$$s'(w) = -\lambda \int \frac{x}{w^2}\phi'\left(\frac{x}{w}\right) f(x)dx \tag{4.30}$$

so that

$$
\begin{aligned}
\int \varepsilon_{sw}(w)R(w)f(w)dw &= \int \frac{ws'(w)}{s(w)} R(w)f(w) \\
&= -\lambda \int\int \frac{x}{w}\phi'\left(\frac{x}{w}\right) f(x)N(w)f(w)dxdw
\end{aligned}
\tag{4.31}
$$

where we have used the steady-state relation $sN = R$. Now, exchanging the roles of x and w in (4.16) and differentiating with respect to x we have

$$R'(x) = \lambda \int \frac{1}{w}\phi'\left(\frac{x}{w}\right) f(w)N(w)dw \tag{4.32}$$

so that (4.31) can be written as

$$\int \varepsilon_{sw}(w)R(w)f(w)dw = -\int xR'(x)f(x)dx = -\int \varepsilon_{Rw}(x)R(x)f(x)dx \tag{4.33}$$

which is (4.17).

Proof of Proposition 4.4

Differentiating (4.19) with respect to w, we have

$$s^{e'}(w) = -\lambda \int \frac{x}{w^2}\phi'\left(\frac{x}{w}\right) f(x)dx \tag{4.34}$$

so that

$$\int \varepsilon^{\mathrm{e}}_{sw}(w)s^{\mathrm{e}}(w)N(w)f(w)dw = -\lambda \int\int \frac{x}{w}\phi'\left(\frac{x}{w}\right)f(x)N(w)f(w)dxdw \tag{4.35}$$

Exchanging the roles of x and w in (4.20) and differentiating with respect to x, we have

$$R^{\mathrm{e}'}(x) = \lambda \int \frac{1}{w}\phi'\left(\frac{x}{w}\right)f(w)N(w)dw \tag{4.36}$$

so that (4.35) can be written as

$$\int \varepsilon^{\mathrm{e}}_{sw}(w)s^{\mathrm{e}}(w)N(w)f(w)dw = -\int xR^{\mathrm{e}'}(x)f(x)dx = -\int \varepsilon^{\mathrm{e}}_{Rw}(x)R^{\mathrm{e}}(x)f(x)dx \tag{4.37}$$

Now, in the economy as a whole (but not firm by firm), total recruits from employment must equal total separations to employment which means that $\int s^{\mathrm{e}}(w)N(w)f(w)dw = \int R^{\mathrm{e}}(w)f(w)dw$. Dividing both sides of (4.37) by this leads to (4.21).

Proof of Proposition 4.5

Rearranging (4.22), we have

$$R^{\mathrm{n}} = \frac{1-\theta_{\mathrm{R}}}{\theta_{\mathrm{R}}}R^{\mathrm{e}} \tag{4.38}$$

which can be written as

$$\log(R^{\mathrm{n}}) = \log\,(R^{\mathrm{e}}) + \log\left(\frac{1-\theta_{\mathrm{R}}}{\theta_{\mathrm{R}}}\right) \tag{4.39}$$

Differentiation leads to (4.23).

Proof of Proposition 4.6

Denote the density function of ξ by $\varphi(\xi)$. The survivor function over a length of time τ will now by given by

$$S(w,\tau) = \int \exp(-\xi_w - \beta_\tau)\varphi(\xi)d\xi \tag{4.40}$$

and the estimate of the wage elasticity will be given by the elasticity of $-\log(S(w,\tau))$ with respect to the wage.

If ξ has a gamma distribution, then (4.40) can be written as

$$S(w,\tau) = \frac{\mu/\sigma^2}{\Gamma(\theta)}\int \exp(-\xi_w - \beta_\tau)\exp\left(-\xi\frac{\mu}{\sigma^2}\right)\left(\frac{\mu\xi}{\sigma^2}\right)^{(\mu/\sigma)^2} d\xi \tag{4.41}$$

After some rearrangement, this can be written as

$$S(w,\tau) = \left(\frac{(\mu/\sigma^2)}{w^{-\beta}\tau + (\mu/\sigma^2)}\right)^{\xi} \frac{1}{\Gamma(\xi)} \int \exp\left(-\xi_w - \beta_\tau + \frac{\mu}{\sigma^2}\right)$$

$$\times\left(w^{-\beta_\tau} + \left(\frac{\mu}{\sigma^2}\right)\right)^{(\mu/\sigma)^2} \xi^{(\mu/\sigma)^2-1} d\xi$$

$$= \left(\frac{(\mu/\sigma^2)}{w^{-\beta_\tau} + (\mu/\sigma^2)}\right)^{(\mu/\sigma)^2} \tag{4.42}$$

Taking logs, we have

$$(w,\tau) = \left(\frac{\mu}{\sigma}\right)^2 \left(\log(\mu) - \log(\mu + \sigma^2 w^{-\beta_\tau})\right) \tag{4.43}$$

Taking the elasticity of $-\log(S)$ with respect to the wage leads to

$$\frac{\partial \log(-(w,\tau))}{\partial \log(w)} = -\beta \frac{(\sigma^2/\mu)w^{-\beta_\tau}}{1 + (\sigma^2/\mu)w^{-\beta_\tau}} \frac{1}{\log(1 + (\sigma^2/\mu)w^{-\beta_\tau})} > -\beta \tag{4.44}$$

(4.44) shows that the estimated wage elasticity is biased towards zero by the presence of unobserved heterogeneity. The size of the bias is increasing in τ so the bias will be lower when a shorter period is used for estimation.

Appendix 4B

This appendix considers two generalizations of the static model of section 4.1.

The Employer Size–Wage Effect and Dynamic Labor Supply Curves

For the most part, our regressions have been of the current wage on current employment. As there is likely to be a difference between the short-run and long-run elasticity of the labor supply curve (see the discussion in section 2.2), one might wonder which elasticity is estimated using cross-sectional data when the true labor supply curve is dynamic.

Suppose the dynamic labor supply curve can be written in the following log-linear isoelastic form:[11]

[11] This might come from the equation $N_t - N_{t-1} = R(w_t) - s(w_t)N_{t-1}$ which says that the change in employment is the difference between recruits and quits.

$$w_t = \varepsilon^s(n_t - n_{t-1}) + \varepsilon n_{t-1} + \nu_{wt} \tag{4.45}$$

where ε^s is the short-run elasticity and ε the long-run elasticity. When one estimates a static regression of w_t on n_t one will estimate

$$E(w_t \mid n_t) = \varepsilon^s n_t + E((\varepsilon - \varepsilon^s)n_{t-1} + \nu_{wt} \mid n_t) \tag{4.46}$$

To work out the last term one needs to know the correlation between n_t and n_{t-1}. A simple model is the following:

$$n_t = \beta n_{t-1} + \nu_{nt} \tag{4.47}$$

where β is a measure of the persistence in employment. One should think of this as being a reduced-form equation for employment. We will assume that $\nu_t = (\nu_{wt}, \nu_{nt})$ is independent of n_{t-1} and jointly normally distributed with mean zero and covariance matrix Σ. Denote by σ_w^2 the variance of ν_w, σ_n^2 the variance of ν_n and σ_{wn} the covariance between ν_w and ν_n. Given these assumptions, the unconditional distribution of n_t (and n_{t-1}) will be normal with variance $\sigma_n^2/(1 - \beta^2)$ and $\mathrm{Cov}(n_t, \nu_{nt}) = \sigma_n^2$. Hence, we will have

$$E(\nu_{wt} \mid \nu_{nt}) = \frac{\sigma_{wn}}{\sigma_n^2}\nu_{nt}$$

$$E(\nu_{nt} \mid n_t) = (1 - \beta^2)n_t \tag{4.48}$$

Putting these into (4.48) leads to

$$E(w_t \mid n_t) = \left(\beta\varepsilon + (1 - \beta)\varepsilon^s + \frac{\sigma_{wn}(1 - \beta^2)}{\sigma_n^2}\right)n_t \tag{4.49}$$

The last term is the simultaneous equations bias caused by the potential correlation between the errors in wage and employment equation: this term could be eliminated by the use of suitable instruments. The other term shows that the estimated elasticity will be a weighted average of the short- and long-run elasticities with the weight being determined by the persistence in employment. So, if employment has no persistence, we will estimate the short-run supply curve and if it has full hysteresis, then we will estimate the long-run elasticity. As the evolution of employment within plants seems quite close to a random walk, it is likely that the cross-sectional correlation between wages and employment estimates the long-run elasticity.

The Employer Size–Wage Effect and the Labor Cost Function

Section 2.3 introduced the generalized model of monopsony and recommended the use of the labor cost function to think about the extent of monopsony in the labor market. Yet, section 4.1 has reverted to a simple monopsony model in which the wage is the only instrument available to the employer for influencing its supply of labor. In this section, we show that the conclusions of the previous section are robust to using the labor cost function approach. Recall that the labor cost function $C(w, N)$ gave the per worker costs of recruitment and training if the firm pays a wage w and wants to have employment of N. To capture this idea assume that, if the firm spends C per worker on recruitment/training activities, its labor supply curve, (4.2), is modified to become

$$w = BC^{-\gamma}N^{\varepsilon} \tag{4.50}$$

where the isoelastic functional form is chosen for convenience. The formula for the rate of exploitation needs to be modified for the presence of C: the natural measure to use is $[Y' - w - C]/[w + C]$ as workers should not expect to receive their costs of training and recruitment.

As C is likely to be unobserved by the econometrician, one might think that the presence of C makes it very difficult to estimate the rate of exploitation. However, the following proposition shows that, once one has appropriately modified the formula for the rate of exploitation, the unobservability of C causes no problems and an estimate of the ESWE gives us the correct parameter estimate.

Proposition 4.8. *The rate of exploitation is given by*

$$\frac{Y'(N) - (w + C)}{w + C} = \frac{\varepsilon}{1 + \gamma} \tag{4.51}$$

and the"reduced-form" labor supply curve after concentrating out the optimal choice of C is given by

$$w = \gamma^{-\gamma/(1+\gamma)}B^{1/(1+\gamma)}N^{\varepsilon/(1+\gamma)} \tag{4.52}$$

Proof. Given N, C will be chosen to minimize $(w + C)$ which, using (4.50), leads to the first-order condition

$$1 = \gamma BC^{-(\gamma+1)}N^{\varepsilon} \quad \Rightarrow \quad C = \gamma w \tag{4.53}$$

Substituting this expression for C into (4.50) and rearranging leads to

$$w = \gamma^{-\gamma/(1+\gamma)}B^{1/(1+\gamma)}N^{\varepsilon/(1+\gamma)} \tag{4.54}$$

which is (4.52). N will then be chosen to maximize $Y(N) - (1 + \gamma)wN$ where w is given in (4.54). Using the fact that (4.53) implies that $(w + C) = (1 + \gamma)w$, leads to the first-order condition of (4.51).

(4.51) says that we want to be able to estimate $\varepsilon/(1 + \gamma)$ to estimate the rate of exploitation while (4.52) says that it is exactly the parameter we would expect to estimate if we run a regression of $\log(w)$ on $\log(N)$. Of course, all the problems we have discussed earlier surrounding the estimation of (4.52) still apply: it is just that acknowledging the labor cost function causes no additional problem.