SERIE DOCUMENTOS DE TRABAJO

FACULTAD ECONOMÍA Y NEGOCIOS
UNIVERSIDAD DE CHILE

DEPARTAMENTO DE ECONOMÍA

SDT 332

THE UNION EFFECT ON WAGES IN CHILE: A TWO-STAGE APPROACH USING PANEL DATA

Autores: Oscar Landerretche, Nicolás Lillo, Esteban Puentes

Santiago, Febrero de 2011

The Union Effect on Wages in Chile: A Two-Stage Approach Using Panel Data

Oscar Landerretche Nicolás Lillo Esteban Puentes*

January 2011

Abstract

We estimate the wage premium in Chile for the period 2004-2009 following a two-stage procedure that corrects the endogeneity of the union decision process. We find that there is true state dependence in the union decision. The wage premium after controlling for endogeneity is close to 20%. At the same time we find evidence that unions tend to increase more the wages of the low end of the wage distribution. Economic sectors are important in the wage equation but seem less relevant in explaining union membership.The most relevant factor that explain union membership is firm size.

*We gratefully acknowledge the funding of Iniciativa Científica Milenio to the Centro de Microdatos (Project P07S-023-F). We thank José Miguel Benavente for his valuable comments and suggestions. We also thank participants at the Labor Economics Seminar at the LACEA 2010 Conference

1 Introduction

This paper adds to the robust empirical literature on union wage premiums[1]. Most of the literature on the individual level union effect on wages uses cross-section data. In that context, there have been two main approaches to solve the endogeneity of union status: IV methods and two-step sample-selection correction methods. The former has been criticized due to the fact that it can't be used to sign the direction of the selection process since IV imposes opposite signs on the selection terms (Robinson (1989*b*)). The latter approach, beginning with Lee (1978), does not fall into such pitfalls but has been criticized due to the instability of the parameter estimates across studies.

A second group of work on the union wage differential question is based on longitudinal data. As Robinson (1989*a*) notes, this group of work appears to obtain more stable but smaller estimates of the union wage premium. However, longitudinal estimates can be affected by measurement error[2]. Longitudinal estimation of the union wage effect has traditionally controlled for individual heterogeneity and its correlation with union status through a fixed-effects approach (see Jakubson (1991)).

In this paper we attempt to execute a well identified estimation of the union premium, that is, the increase in wages associated to union membership for Chile. To do this we use a two-step procedure on panel data detailed in Vella & Verbeek (1999) and applied for the union question in Vella & Verbeek (1998). The procedure first estimates the probability of union membership and then corrects the wage equations using endogeneity correction terms calculated with the first stage estimates. An initial conditions problem arises due to the use of the lagged union status of an individual as an exclusion variable for the wage equations (see Heckman (1981*b*)). To solve the initial conditions problem we use the Conditional Maximum Likelihood estimator introduced by Wooldridge (2005) that also corrects other sources of

[1]For a relatively recent survey of the literature see Hirsch (2004).

[2]Card (1996) exploits this problem to identify the union wage effect in a longitudinal setting using a verification survey.

endogeneity in the first stage (see Arulampalam & Stewart (2009)).

Our estimates are based on two assumptions and two samples. The two assumptions is wether unions also affect the returns on observable characteristics or not, and the two samples are based on occupational categories. We find that our results do not vary too much across the two assumptions for the full sample, but they do have a significant difference when looking at the restricted sample. Our preferred estimation results indicate a union wage premium of 22-23%.

Furthermore, the first stage estimates yield insight into the type of state dependence in union status determination. In the language of Heckman (1981*a*), we find evidence of both *true state dependence* and *spurious state dependence* in the determination of the union status of wages. That is, both the lagged union status and the individual specific unobservable component in the union membership equation are responsible for the high autocorrelation in union status across time. With respect to the sorting process, we test two hypothesis: hierarchical sorting and comparative advantage sorting. We find evidence of the latter.

We must stress that this paper does not aim to evaluate the effects of unionization on the labor market as a whole, since our estimates are focused on individual-level union wage effects. The effect of unionization on employment as a whole and at the individual level are not studied in this paper, but is certainly a very important topic when evaluating different policy proposals. More research is needed in that area and in particular for developing countries.

This paper is organized as follows. The next section introduces the roll of unions in the Chilean labor market. Then we take a look at the econometric framework. Section 4 discusses the estimation procedure. Section 5 then describes the database and discusses the variables to be included in our estimation and some descriptive statistics. In section 6 we present the results and discuss them. Finally section 7 concludes.

2 Unions and Collective Negotiation in the Chilean Labor Market

Chile is known around the World for being a market friendly economy that adopted very early most of the recommendations of the Washington Consensus. The labor market is no exception. During the 1993-2009 period, average labor productivity in Chile grew in 2.6 percent on average while real wages grew 2.8 percent. Of course, these numbers include some significant shocks and structural transformations that this economy has been subject to during the 1990's. Still if one tries to isolate a more recent and crisis free period (2001-2008) average labor productivity grew 1.66 percent, while real wages grew 1.70 percent.[3] Hence, on average, the Chilean labor market behaves as a competitive labor market should. This is not as so surprising if one looks at survey based qualitative assessments of the Chilean labor market such as the World Competitiveness Report or the Economic Freedom Index which usually locate Chile among the higher third in flexibility of the labor market. It is not surprising to see average wages moving roughly with productivity.

This does not mean that Chile has not had some important transformations in the labor market during the last decades. Chile has been characterized for a long time as being a country with a resiliently low labor participation rate of around 56 percent (it was 53 percent 25 years ago), which is a result of a low but rising female labor participation rate of 42 percent (26 percent 25 years ago) and a falling male labor participation rate of 70 percent (75 percent 25 years ago). In the case of male workers this is associated to the rapid increase in tertiary education so it is no cause for alarm, but the local common sense is that female participation is too low and is increasing too slow. Of those employed, around 68 percent represent dependant workers (and hence plausibly unionized). This number is interesting because it had been falling from levels around 65 percent two decades ago to around 63 percent

[3]These calculations are based on public data from the Central Bank and the National Institute of Statistics.

in 2005. Since then there has been an increasing trend in labor dependency which peaked just before the sub-prime crisis at 70 percent. This trend is a result of the passage of labor bills that made outsourcing more expensive for firms.

On the other hand, unionization rate trends in Chile are not that exceptional. Depending on the data source and the base used (with or without civil servants) to calculate the rate, Chilean unionization rates are in the 12-14 percent range. These rates have been falling steadily during the last decades, as is the case of most western economies. In the case of Chile they were in the 18-20 percent range in the early 1990s. These numbers are comparable to what one sees in economies such as the US (with which we compare our results through the work of Vella & Verbeek (1998)) and Spain, but clearly lower than the rates one observes in more developed European countries which are closer to 30 percent (with Scandinavian countries reaching well over 50). The fall in unionization rates in Chile are very similar to those observed in the US during the last twenty years and slightly lower than those in Europe (for a comparison of trends see Visser (2006)).

However, Chile does have some peculiarities which are interesting to point out. One is that the local Labor Law does not make union membership a compulsory requisite to benefit from the results of collective negotiation. A non unionized worker that wishes to receive the same benefits that were negotiated by the union only has to pay 75% of the fee and is automatically awarded the benefits. This, of course, incentives free rider behavior, so in reality the percentage of workers that are covered by the collective negotiations of unions is slightly higher, although there are not good estimations of this effect. On the other hand, the Chilean labor Law allow for an instrumental para-union negotiation structure called the "negotiation group" to enter into collective negotiations. On the other hand, not all unions do collective negotiation every year and some unions never negotiate. Official government estimates are that slightly less than 10% of workers are involved or affected by some sort of collective negotiation. This number was closer to 15 percent two decades ago, but has stopped falling for a couple of years (see Costa & Mizala (2008)).

There is substantial heterogeneity of unionization across sectors. Unsurprisingly these are Chile's main extractive export industry: mining, utilities, and two service sectors: health and education. In these four sectors, around 22 percent of firms have unions. Next are financial services with 12 percent, and non metallic manufactures (mainly foodstuffs) with 10 percent. Around 54 percent of large firms have unions; this falls to 22 percent in medium firms, 3 percent for small, less than 1 percent for micro. These categories are by sales levels, however one must bare in mind that the Chilean Labor Law only allows unions in firms with more than 8 workers, so part of this is legally induced (for more details see *Resultados de la Sexta Encuesta Laboral ENCLA 2008* (2009)).

Hence, in a sense, the Chilean labor market is not necessarily a very exceptional one. It probably reflects most global trends in unionization rates and on average works very similarly to a competitive market.

3 Econometric Framework

The wage of an individual in a specific sector at a certain period in time is determined by a set of observable and unobservable characteristics of both the individual and the firm. We shall model the unobservable characteristics in each wage equation as the sum of two random components: an individual-specific fixed effect and an individual/time-specific random effect. Hence, the wage equation for sector $j = 0,\ 1$ can be expressed as

$$w_{j,it} = \beta_j' X_{it} + \alpha_{j,i} + \epsilon_{j,it} \qquad t = 1, \ldots, T;\ i = 1, \ldots, N; \tag{3.1}$$

where $w_{j,it}$ denotes the wage of individual i in sector j at time t. Likewise, X_{it} is a K-vector of observable characteristics and β_j is a K-vector of unknown parameters that describe the effect of these observable characteristics on wages in sector j. Finally, $\alpha_{j,i}$ is an individual random effect in sector j and $\epsilon_{j,it}$ is a random effects specific to sector j of individual i at time t.

We shall model the sorting process of individuals into the union and non-union sector across time by a discrete-data first order Markov process. We shall make union status depend on a series of observable exogenous variables and other unobservable characteristics. As in the wage equations, the unobservable characteristics will be decomposed into an individual specific and an individual/time specific component. The exogenous variables shall include a series of individual and firm specific characteristics. We shall also include the lagged union status as an explanatory variable[4]. Henceforth union status shall be described by the following process:

$$y_{it} = \gamma' Z_{it} + \delta \cdot d_{it-1} + \theta_i + \eta_{it} \qquad t = 1, \ldots, T;\ i = 1, \ldots, N \tag{3.2}$$

$$d_{it} = \begin{cases} 1 & \text{if } y_{it} \geq 0, \\ 0 & \text{if } y_{it} < 0; \end{cases}$$

where y_{it} is an unobserved latent variable which governs the union status of a worker, d_{it} is the union status of the worker ($d_{it} = j$ if individual i is in state j at time t), Z_{it} is the vector of exogenous variables, γ its coefficient vector, and δ the coefficient on the interactions with the lagged union status d_{it-1}. θ_i is the individual effect and η_{it} is the individual/time-specific effect. Note that since Z_{it} includes an intercept the lagged union status has a level effect on y_{it}.

The inclusion of the lagged union status allows us to test if there's *true state dependence* or *spurious state dependence*. In the language of Heckman (1981*a*), the former refers to state dependence in which past states actually influence future decisions while the latter is a form of state dependence that occurs because there are certain individual characteristics that influence a decision and do not change over time. In both cases there's a significant amount of autocorrelation in an individuals decision, but the source of that correlation can have very different explanations depending of the type of state dependence.

True state dependence in the union status decision can be caused by non-pecuniary

[4]The inclusion of the lagged union status will lead us to the initial conditions problem, which we will address below.

benefits of prolonged union membership. Moreover, union membership can provide more labor stability and thus past union membership can increase the chances that a worker either doesn't change jobs or doesn't lose it altogether. This in turn increases the chances he/she is unionized the next period.

The distribution of the individual specific and individual/time specific unobservable components (in both wage *and* union status equations) are assumed to be independent across individuals. Following Vella & Verbeek (1999), denote $v_{it} = \theta_i + \eta_{it}$, $u_{j,it} = \alpha_{j,i} + \epsilon_{j,it}$, v_i as the T-vector of v_{it}'s, and $u_{j,i}$ as the T-vector of $u_{j,it}$'s. We assume the following error structure:

$$v_i \,|_{X_i} \sim i.i.d.\, N\left(0, \sigma_\theta^2 \iota\iota' + \sigma_\eta^2 I_T\right), \tag{3.3}$$

$$E\left\{u_{j,it} \,|X_i, v_i\right\} = \tau_{1j} v_{it} + \tau_{2j} \bar{v}_i. \tag{3.4}$$

where X_i is a $K \times T$-matrix of observed characteristics, ι is a T-vector of ones, I_T is the identity matrix of size T, τ_{1j} and τ_{2j} are unknown parameters, and $\bar{v}_i = (1/T)\sum_{t=1}^{T} v_{it}$. Equation (3.3) imposes normality on the unobservables in the union status equation and that they have a strict error components structure which excludes autocorrelation in η_{it}. Therefore all dynamic effects are captured by the lagged union status and its interactions in equation (3.2). Equation (3.4) allows the ϵ_{it} to be heteroscedastic and autocorrelated but imposes strict exogeneity of X_{it}.

As we have said, the endogeneity problem we have to tackle arises from the possible correlation between the unobservable characteristics in the wage and union status equations. Specifically, we allow the unobserved characteristics in equations (3.1) and (3.2) to be correlated through the individual specific effects and through the individual/time specific random effects. Following Vella & Verbeek (1998), we allow for four covariances ($\sigma_{j,\alpha\tau}$ and $\sigma_{j,\epsilon\eta}$ for $j = 0, 1$) to be non-zero, while the covariance between the unobservable characteristics in the wage equations of each sector is left unspecified. All other covariances are set to zero.

Assuming a trivariate normal distribution for the unobservables, the joint distri-

bution is given by:

$$\begin{bmatrix} u_{1,i} \\ u_{0,i} \\ v_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{1,\alpha} \iota\iota' + \sigma^2_{1,\epsilon} I_T & \text{unspecified} & \sigma_{1,\alpha\theta} \iota\iota' + \sigma_{1,\epsilon\eta} I_T \\ & \sigma^2_{0,\alpha} \iota\iota' + \sigma^2_{0,\epsilon} I_T & \sigma_{0,\alpha\theta} \iota\iota' + \sigma_{0,\epsilon\eta} I_T \\ & & \sigma^2_{\theta} \iota\iota' + \sigma^2_{\eta} I_T \end{bmatrix} \right). \quad (3.5)$$

If $\sigma_{j,\alpha\theta}$ and $\sigma_{j,\epsilon\eta}$ for $j = 0, 1$ are non-zero, then we have an endogenous-switching regime. The sign of these correlations will determine if the type of sorting, whether it's *hierarchical* or *comparative-advantage* sorting. In the spirit of the Roy (1951) model, if we interpret unobservables as skill endowment, then the first type of sorting means that workers which select into the union sector will be those with higher skills[5] and would receive higher relative wages in whatever sector randomness chooses to put them. On the other hand, comparative-advantage sorting means that skills are specific to a sector, so what makes an individual select into the union sector only raises his wage in that sector (i.e., workers perform differently in the two sectors and sort accordingly).

To see the point about sorting more clearly, suppose $\sigma_{j,\epsilon\eta} = 0 \, \forall j$. *Hierarchical* sorting occurs when $\sigma_{1,\alpha\theta} > \sigma_{0,\alpha\theta} > 0$. That is, on average, union workers will have higher α_j's. On the other hand, *comparative-advantage* sorting requires that $\sigma_{1,\alpha\theta} > 0$ and $\sigma_{0,\alpha\theta} < 0$, so union workers will have relatively higher α_1's but relatively lower α_0's and vice-versa.

4 Estimation Procedure

Given the fact that workers are observed only in the union status that they have at any point in time, estimating the parameters in the wage equations using simple OLS will give biased and inconsistent estimates. A correction for the selection bias is required. Restricting returns on the exogenous variables to be identical to each

[5]The distribution of the unobservable characteristics is such that, on average, workers in the union sector have positive values of the unobservables while the workers in the non-union sector have negative values of the unobservables.

sector, we obtain the following wage equation:

$$w_{it} = \beta' X_{it} + \mu \cdot d_{it} + e_{it} \tag{4.1}$$

where $e_{it} = d_{it}(\alpha_{1,i} + \epsilon_{1,it}) + (1 - d_{it})(\alpha_{0,i} + \epsilon_{0,it})$, and μ captures the union effect on wages.

Conditioning on the union history of an individual, the linear projection of the wage equation is

$$E\left\{w_{it}\,|Z_i, D_i\right\} = \beta' X_{it} + E\left\{d_{it}\,|Z_i, D_i\right\} + E\left\{e_{it}\,|Z_i, D_i\right\}, \tag{4.2}$$

where D_i denotes the T-vector of union status history of the individual. The last term on the right hand side of equation (4.2) cannot be assumed to be zero valued and therefore we use a two-step control function approach to correct the bias. The first stage consists of estimating the parameters in (3.2) by maximum likelihood. In the second stage, the wage equations are corrected for selection bias using endogeneity correction terms evaluated at the parameter estimates obtained in the first stage.

4.1 The First Stage Likelihood Function

To derive the first stage likelihood function, we first follow Heckman (1981*b*). The transition probability that individual i at time t is in a union given θ_i is given by

$$\Pr\left[d_{it}\,|d_{it-1}, \theta_i\right] = \Phi\left\{\left[\gamma' Z_{it} + \delta \cdot d_{it-1} + \theta_i\right] \cdot \left[2d_{it} - 1\right]\right\},$$

where $\Phi\{\cdot\}$ is the univariate normal cumulative distribution function (which comes from the normality assumption made in equation (3.5)). Thus, the likelihood function for the full sample of T observations per person, given the non-stochastic union status

d_{i0}, is given by

$$\mathcal{L} = \prod_{i=1}^{N} \int_{-\infty}^{\infty} \prod_{t=1}^{T} \Phi \left\{ [\gamma' Z_{it} + \delta \cdot d_{it-1} + \theta_i] \cdot [2d_{it} - 1] \right\} f(\theta) \, d\theta,$$

where $f(\theta)$ is the density function of θ and it need not be normal (although we have already assumed so in (3.5)).

Usually when panel data is used, we observe d_{it} from $t = 1$ onwards, and as such cannot observe d_{i0}, the lagged union status at $t = 1$. Estimation could begin at $t = 2$ and take d_{i1} as exogenous, but since by construction d_{i1} is correlated with the individual specific unobserved component, estimates will be biased and inconsistent. This is what Heckman (1981*b*) called the "initial conditions" problem.

Heckman provided a solution based on an approximation of the first period marginal density using pre-sample variables. This procedure can be implemented in STATA©using Stewart (2006)'s -redprob- command, but is computationally prohibitive and unstable. An alternative two-step procedure in the spirit of Heckman's is Orme (2001)'s procedure. However, Orme's method requires the assumption of low correlation between the first period union status and the individual specific unobserved effect, which is too strong in our context of a short panel.

We use a method proposed by Wooldridge (2005) which is computationally accessible and imposes no requirements on the data. Additionally, the Wooldridge method allows individual effects to be correlated with explanatory variables, which partly control for the endogeneity between the explanatory variables and union status. The details of the procedure are explained in appendix A.

4.2 The Second Stage and Endogeneity Correction Terms

Continuing with the case of homogeneous returns on observable characteristics, the last term in (4.2) can be expressed as

$$E\left\{e_{it}\left|Z_i, D_i\right.\right\} = E\left\{\alpha_{j,i}\left|Z_i, D_i\right.\right\} + E\left\{\epsilon_{j,it}\left|Z_i, D_i\right.\right\}. \tag{4.3}$$

Using the standard formulae for the conditional expectation of normally distributed vectors, we express the two terms on the RHS of (4.3) as:

$$E\left\{\alpha_{j,i}\left|Z_i, D_i\right.\right\} = \sigma_{j,\alpha\theta}\left\{\frac{T}{\sigma_\eta^2 + T\sigma_\theta^2} E\left\{\bar{v}_i\left|Z_i, D_i\right.\right\}\right\} = \sigma_{j,\alpha\theta}\mathcal{C}_i \tag{4.4}$$

$$E\left\{\epsilon_{j,it}\left|Z_i, D_i\right.\right\} = \sigma_{j,\epsilon\eta}\left\{\frac{E\left\{v_{it}\left|Z_i, D_i\right.\right\}}{\sigma_\eta^2} - \frac{T\sigma_\theta^2}{\sigma_\eta^2(\sigma_\eta^2 + T\sigma_\theta^2)} E\left\{\bar{v}_i\left|Z_i, D_i\right.\right\}\right\} = \sigma_{j,\epsilon\eta}\mathcal{C}_{it} \tag{4.5}$$

In the second stage, once the correction terms $\mathcal{C}_i$ and $\mathcal{C}_{it}$ are calculated[6], they are included as additional variables to equations (3.1). Using simple GLS we can obtain consistent estimates of $\sigma_{j,\alpha\theta}$ and $\sigma_{j,\epsilon\eta}$ in addition to β.

To achieve identification in the second stage, we omit from the wage equations the lagged union status dummy variable. This rests on the assumption that lagged union status is weakly exogenous given its predetermination. As argued by Vella & Verbeek (1998), lagged union status captures movement costs of changing jobs between the union and non union sector, then it affects current status, but not current wages. Also, if the benefits of being in a union are small, lagged union status should not affect importantly actual wages, and if the benefits were large, we would observe queues to join unions. Also, there is a correlation between lagged and actual wages through the individual effects, the correlation disappears once we control for individual specific effects.

In the case of heterogeneous returns on observable characteristics we shall esti-

[6]To calculate $\mathcal{C}_i$ and $\mathcal{C}_{it}$ we follow the approach in Vella & Verbeek (1999) which is detailed in appendix B.

mate the wage equations for the union and non-union sectors separately and as such obtain consistent estimates of β_1 and β_0.

5 Data and Variables

In this section we describe the data source, how the sample was constructed, the variables used and present some descriptive statistics.

5.1 Data and Sample Construction

The data used comes from the Social Protection Survey (Encuesta de Protección Social) taken by the Microdata Center at the University of Chile in cooperation with the University of Pennsylvania. There are so far four editions of the survey; in 2002, 2004, 2006, and 2009. The 2002 survey was focused on individuals that were covered by either the private or public pension system, and asked them to describe their labor stories from 1980 to 2002. In 2004 they were asked to report labor stories again, but for the 2002 - 2004 period, in the 2006 survey for the 2004-2006 period, and in 2009 for the 2006-2009 period. In 2004 new interviewees were added to make the sample representative of the labor force[7], while others who were interviewed in 2002 were skipped and were re-incorporated in the 2006 survey. In 2009 all interviewees had been surveyed in the 2006 edition, but the there was an approximate 12% non-response proportion. The total sample consists of approximately 16,500 interviewees for the 2002 - 2006 editions and 14,500 for the 2009 edition[8].

A labor story contains starting and ending dates, information on employment status, union status, and firm and industry characteristics. We extracted the last labor story from each edition of the EPS to construct a four period balanced panel: 2002, 2004, 2006 and 2009. All individuals who were unemployed or inactive at any of these points in time were dropped from the sample. In each edition of the

[7] In 2002 the survey was representative only of individuals affiliated to the pension system.
[8] More detalis are available in Bravo et al. (2010).

EPS individuals were also asked individual information such as age, sex, education, marriage status among other variables.

The sample was further restricted to waged workers, both in the public and private sectors. This means that workers in the armed forces, entrepreneurs, self-employed, and domestic service workers were dropped from the sample. We employ our estimation technique for two samples; one with all occupation categories and one with restricting occupations to the medium skill type. The first sample consists of 1,154 individuals and the second sample has 635 individuals.

There are a series of sample-selection issues in the way we construct the sample. First of all, we restrict the sample to individuals who weren't unemployed or inactive. A more complete model would include the employment status of an individual as a first stage and the union status in a second stage. If the unobservable components in each stage are somehow correlated, then the sample selection bias grows with the magnitude of that correlation. Second, by restricting the sample to individuals who worked as waged workers, we exclude those who engaged in activities described as "self-employed", which in Chile is synonymous with informality. Informality and unionization may be related through skill, as workers with a lower skill endowment may end up in a "self-employed" status, but if they somehow obtain a waged job they will be more likely to unionize to hold on to that job.

The two main sources of data loss are the above restrictions: no unemployment spells and when working, the individual must be doing so as a waged worker. This excludes workers who for example worked as waged workers in all but one edition of the EPS.

5.2 Variables

From a theoretical perspective, the variables that are usually used as wage determinants will also affect union status. Industries which engage in activities that yield high rents could be more likely to generate unions. In Chile this is specially the case in high fixed-costs industries like mining and utilities. We therefore included

industry dummy variables in the set of explanatory variables. Also, larger firms can have significantly higher union participation due to economies of scales in union management as well as the fact that union bargaining power could increase with firm size, so we include firm size dummy variables as well[9].

Education may also be simultaneously affecting union membership and wages. The correlation between membership and wages can be positive if low education workers might find it harder to find unionized jobs because employers will sort out these workers from the queue when hiring (see Abowd & Farber (1982)). Also more educated workers have more and better information about the benefits of a union and more educated workers can have more resources and are able to wait longer if a negotiation takes time. However, there could be a negative relationship between education and unions if more educated workers change jobs more rapidly and do not want to engage in long term relationships with one firm. Therefore, a priori the effects of education on wages are not unambiguous and probably have significant heterogeneity. We include education as dummy variables that describe three categories: less than 12 years of education, 12 years exactly, and over 12 years of education.

The positive effect of experience on wages has been amply reported, as well as the effect of experience on unionization. Theoretically, less experienced workers can be sorted out of the queue for union jobs and therefore unionization will increase with experience. Yet, this effect may be nonlinear if more specific skills tend to materialize in the latter stages of experience building. We construct experience as potential experience[10].

Individual characteristics such as sex and health limitations can have certain impact on unionization and wages simultaneously. Individual preferences regarding union status can be shaped by these variables. Unionization has been perceived as mostly a male activity, but the correlation hides other factors such as economic

[9]Some individuals that could not name the number of employees in their firm were asked if that number fell within some predefined intervals. We therefore code firm size into four categories: 1 to 9 workers, 10 to 49 workers, 50 to 199 workers, and 200 plus workers.

[10]Potential experience was constructed using the standard formula *age-6-years of education*.

activity of the employer. In fact, women could use unions as a mechanism to protect them from employer discrimination, and thus controlling for other factors women should have a higher tendency to unionize and higher wages in the union sector. On the other hand, those women who aren't sorted out of a unionized firm might end up in lower paying jobs than observationally equivalent men. Health limitations could have an ambiguous effect on unionization and wages. If there is significant discrimination then individuals with health limitations might have a higher incentive to unionize, but at the same time unionized firms will try to sort out these individuals from the queue (see Lee (1978)). Without a union to homogenize them, one would expect individuals with health limitations to have lower wages.

Being in the public sector should have a positive impact on the probability of being unionized, and at the same time a higher wage premium if the individual is indeed unionized. In Chile, on the other hand, by law unionization is illegal for public sector employees, but in practice public sector employees do conduct collective negotiations particularly through the National Federation of Public Sector Employees (ANEF). So even though unions aren't permitted *de jure*, the ANEF is a *de facto* public sector union. Nevertheless, because of the restrictions caused by budgetary rules, the public sector has a vast number of workers with short term contracts, which aren't necessarily covered by the ANEF but still tend to be part of the negotiation process through a more political but indirect mechanism. So if a worker has a short term contract with the public sector and is not covered by the ANEF, it could still enjoy a higher wage than an observationally equivalent non-public sector worker.

5.3 Summary Statistics

Table 1 presents pooled summary statistics in our working sample. The first thing to notice is the increasing unionization rate from 2004 to 2009. This is consistent with the non stationary model of union membership that we propose.

The second thing to notice is the high unionization rate of above 20 percent in both samples, which is high compared to other samples. The *Trabajo y Equidad*

survey, taken in 2007, shows a unionization rate of 14.8%. The unrestricted EPS shows that the unionization rate for all labor stories in 2006 was of 15.1%. In explaining this high number we must stress that our constructed samples consist of only waged workers and workers without unemployment spells between 2002 and 2009. We therefore have a constructed bias towards unionized workers on our sample that should be examined in future research.

Table 1 also shows the descriptive statistics of the explanatory variables in our model. The first noticeable fact is that wages are on average higher in the union sector. The difference between the log(wage/hour) in the union and non-union sectors is .27 in the first sample and .20 in the second sample. Second, there's a clear increase in unionization with education. Third, there is a clear positive relationship between union status and firm size. There's also a positive relationship between being a public sector employee and union status. Finally, on average, experience is slightly higher in the unionized sector.

Confirming what was shown in table 1, in figure 1 it can be seen that the wage distribution of the union sector is shifted to the right of the non-union sector. Of course, this shift is highly biased and cannot be taken to mean anything more than the correlation already mentioned, but the size of the shift does tell us there's ample room to research on significant differences between the two groups even after controlling for the endogeneity of union status.

6 Results

In this section we first look at the results of the first stage estimation of the union status of workers. Then we look at the results from the second stage estimation of the impact of unions on wages; first with homogeneous returns on observables and then allowing for heterogeneity. Finally we provide a summary of our findings and discuss our interpretation of the results.

6.1 First Stage: The Union Status of Workers

Tables 2 and 3 show the results for the first stage. We estimate models for two samples: One sample includes workers in all occupation categories and the other only workers with medium skilled occupations. As a benchmark for comparison we report in each table the results of a random effects probit estimation of the union status equation without including the lagged union status as an explanatory variable and the effect of the Wooldridge correction without introducing lagged union status. A series of correlations which we already hinted at in the previous section now gain statistical significance in the "naive" estimation, but are not robust to the Wooldridge correction. In particular, personal characteristics seem not to play a role in union status determination.

The first and most noticeable result is the high parameter and statistical significance of the lagged union status. In turn, the variance and statistical significance of the unobserved individual specific unobserved component is reduced considerably with respect to the benchmark model[11]. This is to be expected as is the increase in log-likelihood. The likelihood-ratio test that the percentage of unobserved variance is due to the individual specific effect is 0 strongly rejects the null at the 5% confidence level.

In the language of Heckman (1981*a*), true state dependence seems to be more important than spurious state dependence. Hence we seem to find that treating workers to union membership randomly in Chile would significantly increase their union membership henceforth, independently of their inherent propensity to unionize.

The second noticeable fact is that the parameter on the union status of the initial period, from the EPS 2002, is very high and statistically significant. This means that the initial conditions problem is very important. This was to be expected due to the

[11]We use ©STATA 10.1's `xtprobit` command to estimate both models. As such, the variance that this procedure presents in the second model is that of the auxiliary individual specific unobservable component, σ_a (see the appendix for details). To obtain $\hat{\sigma}_\theta$ we take the square root of the sum of σ_a^2 and the variance of the linear prediction of the means on the explanatory variables. That is, $\hat{\sigma}_\theta = (\sigma_a^2 + Var(W_i'\hat{\lambda}))^{\frac{1}{2}}$

fact that our panel is very short[12].

Most of the control variables lose significance for both samples as we depart from the naive estimation. The variables that seem robust to sample and estimation procedure are experience and firm size.

6.2 Second Stage

As explained in section 4, once parameter estimates of the union status equation and the individual specific standard deviation are obtained, the endogeneity correction terms are calculated using the procedure detailed in appendix B. The wage equations are then estimated using a random effects GLS procedure including the endogeneity correction terms as explanatory variables.

We employ two types of endogeneity correction. The first type constrains the returns on the observable characteristics to be identical across sectors, and as such the union effect on wages is captured by a union status dummy variable. Following Vella & Verbeek (1998) and Chrysanthou (2008), we explore the hypothesis of hierarchical or comparative-advantage sorting by estimating two models. The first model of hierarchical sorting simply adds to the set of explanatory variables the endogeneity correction terms. Comparative-advantage sorting is tested by interacting the union status variable with the endogeneity correction terms, while constraining the parameters on the other explanatory variables to be the same across sectors. We also add a benchmark GLS estimation of the union effect on wages without taking into account the endogeneity of union status.

The second type of endogeneity correction estimates two equations, one for the union sector sub-sample and another for the non-union sector sub-sample. The objective of this technique is to capture heterogeneity of the returns on the observable characteristics across sectors. That is, we estimate two different vectors β_1 and β_0 instead of a single vector β as in equation (4.1). The union wage effect is then

[12]Only the means on the public sector dummy and the financial services dummy turned out to be significant for the Wooldridge correction.

measured as the difference between the predicted wages. We therefore obtain three parameters of interest: the Average Treatment Effect (ATE), the Treatment on the Treated Effect (TT), and the Treatment on the Untreated Effect (TUT) (see Heckman et al. (2001)). Hierarchical and comparative advantage sorting are analyzed by looking directly at the significance and sign of the parameter estimates corresponding to the endogeneity correction terms in each sub-sample estimation.

6.2.1 Estimation of the Wage Equations with Homogeneous Returns on Observable Characteristics Across Sectors

The results of the three estimations under homogeneous returns on observable characteristics are presented in tables 4 and 5. Focus first on the uncorrected GLS estimation in each table. The union effect on wages is estimated at approximately 7.68% for the full sample and 7.84% for the restricted sample.

Now focus on the GLS - Hierarchical estimation. When introducing the endogeneity correction terms $\mathcal{C}_i$ and $\mathcal{C}_{it}$, the parameter on union status increases significantly to 0.147 in the full sample and to 0.182 in the restricted sample, which turns out to be a 15.8 and 19.9% increase in wages respectively. The endogeneity correction terms are negative and significant in the full sample estimation and only the individual/time specific term is significant (and negative) in the restricted sample.

The third estimation model is that of the unrestricted sorting process. By interacting union status with the endogeneity correction terms we can test the accuracy of the hierarchical model. Under unrestricted sorting, the union status parameter is estimated to be 0.203 and 0.205 for the full and restricted samples respectively, which means that the union effect on wages is approximately 22-23%.

The endogeneity correction terms turn out to be significant but not robust to the sample used. For the full sample, three out four terms are significant. Specifically, the individual specific terms are significant with opposing signs, while the individual/time specific term for the non-union sector is also significant and negative. For the restricted sample, the union sector individual specific term is significant and

positive while the non-union sector individual/time specific term is significant and negative.

The rest of the control variables have the expected signs and significance. In fact, parameter estimates of these variables are approximately the same across the three specifications. It's also worth noting that the public sector seems not to play a role at all in the wage equation. This is a feature that is common to all three specifications.

6.2.2 Estimation of the Wage Equations allowing for Heterogeneous Returns on Observable Characteristics Across Sectors

The results for the estimation of the second stage wage equations allowing for heterogeneous returns on observable characteristics are presented in table 6. The first thing to notice is that there are, for both samples, important differences in the parameter estimates of the returns on observable characteristics between the union and non-union sectors. Yet, these differences are not large enough to create a big wedge between the different treatment parameters; the ATE, TT and TUT.

The full sample results with heterogeneous returns on observables shows that the union wage effect is around 22-23%, similar to the effect found when restricting returns on observables to be the same and when using unrestricted sorting. Yet, the same cannot be said for the restricted sample results. In fact, the ATE, TT and TUT fall by around 5-6 percentage points.

6.3 Discussion

The results presented in the preceding section show that our search for heterogeneity in union wage effects is marked by unobserved heterogeneity. The significant estimates on the individual specific correlations between the union status decision and the wage equations with different signs means that there is evidence of comparative sorting.

Table 7 summarizes the relevant parameter estimates and wage effects across methodologies. It can be seen clearly that introducing the endogeneity correction

terms in the homogenous returns approach (first three rows for each sample) increases the wage effect. When we consider the heterogeneous effect, it can be see that for the whole sample (all occupations) the effect increases while for the restricted sample (medium skill occupations) the effect falls compared to the GLS-Unrestricted sorting estimate.

Whilst for the whole sample we find that there is comparative advantage sorting only for the individual specific unobserved effects (opposing signs of $\sigma_{1,\alpha\theta}$ and $\sigma_{0,\alpha\theta}$), we find that for the restricted sample the heterogeneity in covariances of unobservable characteristics is distributed between individual specific and individual/time specific effects.

In table 8 the treatment effects for the different samples are presented. For the whole sample we find that the untreated group has a slightly higher union effect than the treated effect. For the restricted sample we find that the treated have experienced a higher effect than the untreated.

How the effect is distributed across the wage distribution is shown in figure 3. The results show that there is wage compression caused by unions, in the sense that the top earners have less to gain from unionization, *but all wage categories have a positive wage premium*. When looking only at the restricted sample results, it seems the effect is more stable, although the gain falls dramatically for the top wage earners.

7 Conclusions

We estimate the effect on union on wages using panel data for a small open economy that is considered as having a very flexible labor market. We follow a two stage approach that controls for individual heterogeneity in the decision to enter a union and in wages. The methodology follows closely Vella & Verbeek (1998), but the first stage uses the correction suggest by Wooldridge (2005). Additionally, we improve on previous literature by estimating different equations for the union and non-union sector. The estimates show that the data is consistent with a model of individual

heterogeneity.

Our results show that union increase wages by a 20%, which is larger than the 8% found if we do not control for endogeneity. At the same time, the evidence suggest that unions have a higher impact on for individuals with low wages.

We find that union membership is mostly driven by firm size and past union status, with tenure also playing a part. Moreover, the high correlation between the public sector and union status seems to be caused by their common relationship with the individual specific unobserved effect.

Economic sectors do not appear to be important in the union membership equation, but they do affect wages, especially in the not unionized sector.

References

Abowd, J. M. & Farber, H. S. (1982), 'Job Queues and the Union Status of Workers', *Industrial and Labor Relations Review* **35**(3), 354–367.

Arulampalam, W. & Stewart, M. B. (2009), 'Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators', *Oxford Bulletin of Economics and Statistics* **71**(5), 659–681.

Bravo, D., Mukhopadhyay, S. & Todd, P. E. (2010), 'Effects of School Reform on Education and Labor Market Performance Evidence from Chile's Universal Voucher System', *Quantitative Economics* **1**, 47–95.

Card, D. (1996), 'The Effect of Unions on the Structure of Wages: A Longitudinal Analysis', *Econometrica* **64**(4), 957–979.

Chrysanthou, G. M. (2008), Estimating Union Wage Effects in Great Britain During 1991 - 2003.

Costa, R. & Mizala, A. (2008), 'Negociación colectiva y Tasa de Sindicalización'.

Heckman, J. J. (1981*a*), Heterogeneity and State Dependence, *in* S. Rosen, ed., 'Studies in Labor Markets', The University of Chicago Press, Chicago and London, chapter 3, pp. 91–139.

Heckman, J. J. (1981*b*), The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process, *in* C. F. Manski & D. L. McFadden, eds, 'Structural Analysis of Discrete Data and Econometric Applications', The MIT Press, Cambridge, chapter 4, pp. 179–195.

Heckman, J., Tobias, J. L. & Vytlacil, E. (2001), 'Four Parameters of Interest in the Evaluation of Social Programs', *Southern Economic Journal* **68**(2), 210–223.

Hirsch, B. T. (2004), 'Reconsidering Union Wage Effects: Surveying New Evidence on an Old Topic', *Journal of Labor Research* **25**(2), 233–266.

Jakubson, G. (1991), 'Estimation and Testing of the Union Wage Effect Using Panel Data', *The Review of Economic Studies* **58**(5), 971–991.

Lee, L.-F. (1978), 'Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables', *International Economic Review* **19**(2), 415–433.

Orme, C. D. (2001), 'Two-Step Inference in Dynamic Non-Linear Panel Data Models', *mimeo* pp. 1–16.

Resultados de la Sexta Encuesta Laboral ENCLA 2008 (2009), Technical report, Direccion del Trabajo, Santiago, Chile.

Robinson, C. (1989*a*), 'The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models', *Journal of Political Economy* **97**(3), 639 – 667.

Robinson, C. (1989*b*), 'Union Endogeneity and Self-Selection', *Journal of Labor Economics* **7**(1), 106 – 112.

Roy, A. D. (1951), 'Some Thoughts on the Distribution of Earnings', *Oxford Economic Papers* **3**(2), 135–146.

Stewart, M. B. (2006), Maximum Simulated Likelihood Estimation of Random Effects Dynamic Probit Models with Autocorrelated Errors.

Vella, F. & Verbeek, M. (1998), 'Whose Wages do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men', *Journal of Applied Econometrics* **13**, 163–183.

Vella, F. & Verbeek, M. (1999), 'Two-step estimation of panel data models with censored endogenous variables and selection bias', *Journal of Econometrics* **90**, 239–263.

Visser, J. (2006), 'Union Membership Statistics in 24 Countries', *Monthly Labor Review* pp. 38–49.

Wooldridge, J. M. (2005), 'Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity', *Journal of Applied Econometrics* **20**, 39– 54.

A Wooldridge's CML Estimator

The CML estimator used in this paper to address the initial conditions problem assumes that the individual specific unobservable component in the union status equation can be approximated using the following distribution:

$$\theta_i | d_{i0}, W_i \sim N(\lambda_0 + \lambda_1 \cdot d_{i0} + \lambda' W_i, \sigma_a^2) \tag{A.1}$$

where W_i is the row vector of all (nonredundant) explanatory variables in all time periods (or their means). We can therefore replace θ_i in (3.2) with

$$\theta_i = \lambda_0 + \lambda_1 \cdot d_{i0} + \lambda' W_i + a_i \tag{A.2}$$

where $a_i | (d_{i0}, W_i) \sim N(0, \sigma_a^2)$. So now we can say that d_{it} follows a probit model with transition probability

$$\Pr\left[d_{it} | d_{it-1}\right] = \Phi\left[(2 \cdot d_{it} - 1) \cdot (\gamma' Z_{it} + \delta \cdot d_{it-1} + \lambda_0 + \lambda_1 \cdot d_{i0} + \lambda' W_i + a_i)\right]. \tag{A.3}$$

Therefore, Wooldridge's CML estimator turns out to be the standard random effects probit estimator but with d_{i0} and W_i included as additional regressors.

B Calculating the Endogeneity Correction Terms

This appendix details the derivation of the endogeneity correction terms. First, recall that we can write the error terms in the wage and union status equations as

$$u_{j,it} = \alpha_{j,i} + \epsilon_{j,it} \tag{B.1}$$

$$v_{it} = \theta_i + \eta_{it}. \tag{B.2}$$

We need to compute the conditional expectations on the RHS of (4.3), that is, the expectation of $u_{j,it}$ given the union history D_i. Using the normality assumption we

can find the conditional expectation of $u_{j,it}$ given v_i using the standard formulae. Thus we obtain

$$E\left\{\alpha_{j,i}\left|v_i\right.\right\} = \sigma_{j,\alpha\theta}\left[\frac{T_i}{\sigma_\eta^2 + T_i\sigma_\theta^2}\bar{v}_i\right] \tag{B.3}$$

$$E\left\{\epsilon_{j,it}\left|v_i\right.\right\} = \sigma_{j,\epsilon\eta}\left[\frac{1}{\sigma_\eta^2}v_{it} - \frac{T_i\sigma_\theta^2}{\sigma_\eta^2(\sigma_\eta^2 + T_i\sigma_\theta^2)}\bar{v}_i\right]. \tag{B.4}$$

To obtain the conditional expectation given the union history, replace the v_{it}'s in (B.3) and (B.4) by their conditional expectation given D_i.

Now to obtain the conditional expectation of v_{it} given D_i we use definition (B.2) and

$$E\left\{\theta_i + \eta_{it}\left|D_i\right.\right\} = \int_{-\infty}^{+\infty}\left[\theta_i + E\left\{\eta_{it}\left|D_i, \theta_i\right.\right\}\right] f(\theta_i|D_i)\, d\theta_i \tag{B.5}$$

where $E\left\{\eta_{it}\left|D_i, \theta_i\right.\right\} = E\left\{\eta_{it}\left|d_{it}, \theta_i\right.\right\}$ is the usual generalized residual of the probit model given by

$$E\left\{\eta_{it}\left|D_i, \theta_i\right.\right\} = (2d_{it}-1)\sigma_\eta\frac{\phi\left((2d_{it}-1)(\gamma' Z_{it} + \delta' Z_{it}\cdot d_{it-1}\theta_i)/\sigma_\eta\right)}{\Phi\left[(2d_{it}-1)(\gamma' Z(i,t) + \delta' Z_{it}\cdot d_{it-1} + \theta_i)/\sigma_\eta\right]}. \tag{B.6}$$

Now we need $f(\theta_i|D(i))$, which is given by

$$f(\theta_i|D_i) = \frac{\prod_{s=t_{0i}}^{T_i}\Phi\left[(2d_{is}-1)(\gamma' Z_{is} + \delta' Z_{is}\cdot d_{is-1} + \theta_i)/\sigma_\eta\right]\phi\left(\theta_i/\sigma_\theta\right)/\sigma_\theta}{\int\left\{\prod_{s=t_{0i}}^{T_i}\Phi\left[(2d_{is}-1)(\gamma' Z_{is} + \delta' Z_{is}\cdot d_{is-1} + \theta_i)/\sigma_\eta\right]\phi\left(\theta_i/\sigma_\theta\right)/\sigma_\theta\right\} d\theta_i}. \tag{B.7}$$

Finally we plug (B.6) and (B.7) into (B.5). To do so it's necessary first to numerically compute the integrals in (B.7) and again in (B.5).

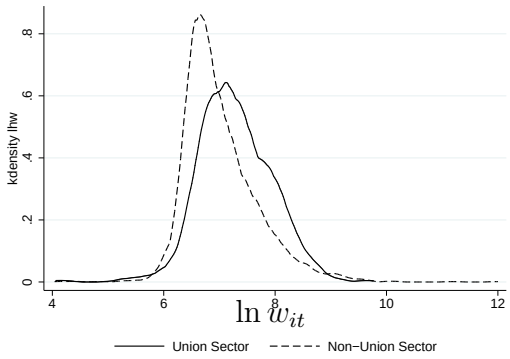
C Tables and Figures

Table 1: Descriptive Statistics

	All Occupations				Medium Skill Occupations			
	Not Unionized		Unionized		Not Unionized		Unionized	
VARIABLES	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Log(Wage/Hour)	7.016	(0.652)	7.286	(0.649)	6.990	(0.532)	7.193	(0.528)
Gender	0.704	(0.456)	0.615	(0.487)	0.690	(0.463)	0.626	(0.484)
Health Limitations	0.0238	(0.153)	0.0313	(0.174)	0.0230	(0.150)	0.0156	(0.124)
Education: 12 years	0.533	(0.499)	0.578	(0.494)	0.688	(0.463)	0.782	(0.414)
Education: >12 years	0.176	(0.381)	0.273	(0.446)	0.110	(0.313)	0.117	(0.322)
Experience	26.71	(11.78)	27.22	(11.04)	25.28	(11.02)	25.40	(10.28)
Experience squared	852.2	(719.5)	862.6	(667.3)	760.5	(632.8)	750.3	(568.3)
Private sector worker	0.875	(0.331)	0.607	(0.489)	0.920	(0.271)	0.700	(0.459)
Ind: Mining	0.0123	(0.110)	0.0313	(0.174)	0.0151	(0.122)	0.0390	(0.194)
Ind: Manufacturing	0.154	(0.361)	0.165	(0.371)	0.211	(0.408)	0.246	(0.431)
Ind: Utilities	0.0100	(0.0995)	0.0162	(0.126)	0.0108	(0.103)	0.0175	(0.131)
Ind: Construction	0.109	(0.312)	0.0441	(0.205)	0.111	(0.315)	0.0351	(0.184)
Ind: Wholesale, Retail, and Hotels	0.185	(0.388)	0.109	(0.312)	0.267	(0.442)	0.148	(0.356)
Ind: Transport, Storage and Comms.	0.0765	(0.266)	0.0557	(0.229)	0.107	(0.309)	0.0565	(0.231)
Ind: Financial Services	0.0781	(0.268)	0.0534	(0.225)	0.0776	(0.268)	0.0721	(0.259)
Ind: Personal Services	0.228	(0.420)	0.465	(0.499)	0.139	(0.346)	0.341	(0.475)
Firm Size: 10 - 49	0.318	(0.466)	0.155	(0.363)	0.305	(0.461)	0.135	(0.342)
Firm Size: 50 - 199	0.219	(0.414)	0.232	(0.422)	0.202	(0.402)	0.185	(0.389)
Firm Size: 200+	0.250	(0.433)	0.574	(0.495)	0.256	(0.437)	0.643	(0.480)
Occ: Professionals	0.0908	(0.287)	0.148	(0.356)				
Occ: Technicians	0.0550	(0.228)	0.103	(0.304)				
Occ: Clerks	0.141	(0.348)	0.151	(0.358)	0.210	(0.408)	0.203	(0.402)
Occ: Service and Retail	0.136	(0.343)	0.150	(0.357)	0.205	(0.404)	0.214	(0.411)
Occ: Skilled Agro and Fish	0.0658	(0.248)	0.0290	(0.168)	0.0503	(0.219)	0.0214	(0.145)
Occ: Craft and Related	0.164	(0.370)	0.115	(0.319)	0.247	(0.431)	0.168	(0.374)
Occ: Plant and Machine Opps.	0.138	(0.345)	0.171	(0.376)	0.208	(0.406)	0.265	(0.442)
Occ: Unskilled	0.198	(0.399)	0.118	(0.323)				
Observations	2,600		862		1,392		513	
# of Individuals	1,154		1,154		635		635	
Unionization Rate 2004				23.5%				20.5%
Unionization Rate 2006				27.4%				25.7%
Unionization Rate 2009				29.9%				28.5%

Figure 1: Kernel Density of Log(Wage/Hour) by Union Status and Sample

All Occupations



Only Medium Skill

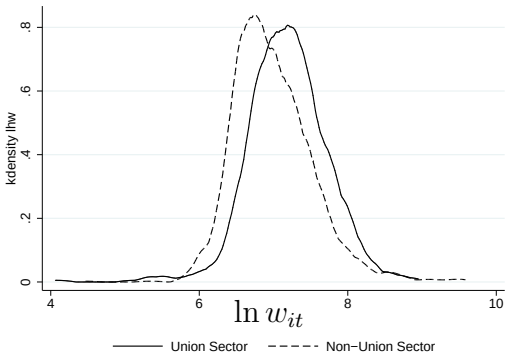
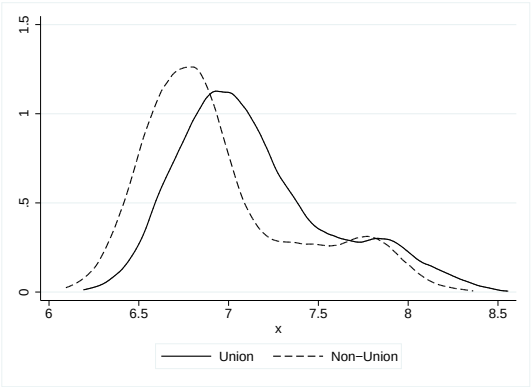


Figure 2: Kernel Distribution of Predicted Wages

All Occupations



Medium Skill Occupations

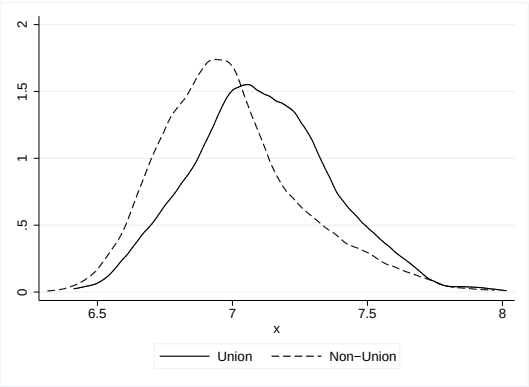


Table 2: Stage 1 Results: All Occupations

VARIABLES	No Correction		CML, No Lag		CML, With Lag	
Lagged Union					0.481***	(0.128)
Gender	-0.0352	(0.130)	0.0545	(0.135)	0.0437	(0.113)
Health Limitations	0.185	(0.244)	0.212	(0.296)	0.158	(0.283)
Education: 12 years	0.446***	(0.166)	-0.142	(0.175)	-0.121	(0.146)
Education: >12 years	0.433*	(0.246)	-0.431	(0.287)	-0.359	(0.240)
Experience	0.0424**	(0.0186)	0.199***	(0.0454)	0.162***	(0.0439)
Experience squared	-0.000425	(0.000306)	-0.00191**	(0.000768)	-0.00168**	(0.000729)
Private sector worker	-0.707***	(0.133)	0.0620	(0.190)	0.101	(0.181)
Ind: Mining	1.038***	(0.344)	0.545	(0.476)	0.547	(0.450)
Ind: Manufacturing	0.527**	(0.212)	0.154	(0.303)	0.157	(0.286)
Ind: Utilities	0.971**	(0.427)	0.664	(0.637)	0.545	(0.602)
Ind: Construction	-0.157	(0.257)	0.140	(0.382)	0.158	(0.359)
Ind: Wholesale, Retail, and Hotels	0.248	(0.225)	0.179	(0.321)	0.180	(0.301)
Ind: Transport, Storage, and Comms.	0.160	(0.256)	-0.271	(0.384)	-0.251	(0.362)
Ind: Financial Services	0.234	(0.265)	0.264	(0.400)	0.278	(0.376)
Ind: Personal Services	1.017***	(0.222)	0.452	(0.343)	0.464	(0.323)
Firm Size: 10 - 49	0.706***	(0.171)	0.489**	(0.211)	0.437**	(0.199)
Firm Size: 50 - 199	1.420***	(0.174)	1.096***	(0.217)	1.008***	(0.204)
Firm Size: 200+	1.929***	(0.169)	1.245***	(0.210)	1.165***	(0.198)
Occ: Professionals	-0.0384	(0.379)	-0.446	(0.460)	-0.448	(0.430)
Occ: Technicians	0.529	(0.391)	0.0952	(0.465)	0.0185	(0.438)
Occ: Clerks	0.119	(0.386)	-0.0639	(0.463)	-0.0637	(0.435)
Occ: Service and Retail	0.363	(0.401)	0.0647	(0.488)	0.104	(0.460)
Occ: Skilled Agro and Fish	0.602	(0.457)	1.199**	(0.582)	1.122**	(0.543)
Occ: Craft and Related	0.451	(0.411)	0.471	(0.516)	0.484	(0.486)
Occ: Plant and Machine Opps.	0.520	(0.408)	0.0396	(0.516)	0.0336	(0.486)
Occ: Unskilled	-0.0126	(0.405)	-0.0620	(0.494)	-0.0206	(0.465)
Union Status 2002			1.469***	(0.146)	1.028***	(0.168)
Constant	-3.746***	(0.589)	-2.700***	(0.876)	-2.422***	(0.736)
Observations	3462		3462		3462	
Individuals	1154		1154		1154	
Log-Likelihood	-1398		-1262		-1255	
σ_a	1.253		1.166		0.880	
χ^2	300.7		229.6		39.29	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Stage 1 Results: Medium Skill Occupations Only

VARIABLES	No Correction		CML, No Lag		CML, With Lag	
Lagged Union					0.579***	(0.183)
Gender	-0.109	(0.184)	-0.200	(0.182)	-0.170	(0.146)
Health Limitations	-0.372	(0.405)	-0.365	(0.502)	-0.349	(0.481)
Education: 12 years	0.304	(0.242)	-0.316	(0.238)	-0.259	(0.192)
Education: >12 years	-0.176	(0.347)	-0.885**	(0.354)	-0.718**	(0.289)
Experience	0.0581**	(0.0287)	0.157**	(0.0644)	0.101	(0.0628)
Experience squared	-0.00100*	(0.000514)	-0.00146	(0.00116)	-0.000903	(0.00111)
Private sector worker	-0.920***	(0.206)	-0.504*	(0.290)	-0.425	(0.274)
Ind: Mining	0.486	(0.463)	-0.396	(0.634)	-0.401	(0.593)
Ind: Manufacturing	0.169	(0.313)	-0.382	(0.451)	-0.403	(0.420)
Ind: Utilities	0.760	(0.586)	0.933	(0.939)	0.634	(0.862)
Ind: Construction	-0.996**	(0.408)	-0.451	(0.637)	-0.432	(0.592)
Ind: Wholesale, Retail, and Hotels	-0.305	(0.341)	-0.480	(0.498)	-0.487	(0.462)
Ind: Transport, Storage, and Comms.	-0.604	(0.372)	-1.171**	(0.555)	-1.164**	(0.514)
Ind: Financial Services	0.126	(0.384)	0.0668	(0.585)	0.0540	(0.542)
Ind: Personal Services	0.680**	(0.346)	0.0592	(0.538)	0.0329	(0.500)
Firm Size: 10 - 49	0.733***	(0.231)	0.338	(0.288)	0.288	(0.267)
Firm Size: 50 - 199	1.473***	(0.239)	0.846***	(0.302)	0.789***	(0.281)
Firm Size: 200+	2.200***	(0.227)	1.141***	(0.281)	1.111***	(0.262)
Occ: Clerks	-0.402*	(0.212)	-0.0869	(0.269)	0.00213	(0.256)
Occ: Service and Retail	-0.0993	(0.233)	0.0632	(0.292)	0.210	(0.282)
Occ: Skilled Agro and Fish	-0.304	(0.450)	1.166	(0.743)	1.115*	(0.674)
Occ: Craft and Related	0.0915	(0.274)	0.580	(0.412)	0.711*	(0.389)
Occ: Plant and Machine Opps.	0.139	(0.261)	0.114	(0.399)	0.185	(0.377)
Union Status 2002			1.903***	(0.199)	1.292***	(0.248)
Constant	-2.510***	(0.654)	-2.571***	(0.890)	-2.239***	(0.722)
Observations	1905		1905		1905	
Individuals	635		635		635	
Log-Likelihood	-756.7		-653.9		-649.1	
σ_a	1.281		1.079		0.742	
χ^2	171.5		97.00		11.71	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Stage 2 Results: All Occupations, Homogenous Returns on Observables

VARIABLES	GLS - No Correction		GLS - Hierarchical		GLS - Unrestricted	
Union	0.0787***	(0.0218)	0.134***	(0.0326)	0.193***	(0.0395)
$\mathcal{C}_i$			-0.426**	(0.172)	-1.010***	(0.258)
Union*$\mathcal{C}_i$					1.099***	(0.373)
$\mathcal{C}_{it}$			-0.0843*	(0.0484)	-0.0675	(0.0604)
Union*$\mathcal{C}_{it}$					0.198	(0.174)
Gender	0.152***	(0.0295)	0.152***	(0.0295)	0.155***	(0.0295)
Health Limitations	-0.0759*	(0.0449)	-0.0750*	(0.0453)	-0.0732	(0.0455)
Education: 12 years	0.329***	(0.0303)	0.318***	(0.0303)	0.312***	(0.0303)
Education: >12 years	0.922***	(0.0576)	0.909***	(0.0575)	0.897***	(0.0577)
Experience	0.0227***	(0.00383)	0.0226***	(0.00384)	0.0262***	(0.00399)
Experience squared	-0.000175***	(6.07e-05)	-0.000165***	(6.03e-05)	-0.000198***	(6.10e-05)
Private sector worker	0.0229	(0.0341)	0.0255	(0.0340)	0.0312	(0.0342)
Ind: Mining	0.322***	(0.0680)	0.313***	(0.0683)	0.319***	(0.0684)
Ind: Manufacturing	0.112***	(0.0306)	0.110***	(0.0306)	0.106***	(0.0305)
Ind: Utilities	0.176**	(0.0711)	0.179**	(0.0713)	0.184***	(0.0713)
Ind: Construction	0.198***	(0.0393)	0.201***	(0.0392)	0.201***	(0.0391)
Ind: Wholesale, Retail, and Hotels	0.110***	(0.0337)	0.104***	(0.0338)	0.0994***	(0.0338)
Ind: Transport, Storage, and Comms.	0.0992**	(0.0468)	0.0967**	(0.0467)	0.0853*	(0.0468)
Ind: Financial Services	0.228***	(0.0432)	0.222***	(0.0430)	0.221***	(0.0429)
Ind: Personal Services	0.0741**	(0.0363)	0.0674*	(0.0364)	0.0663*	(0.0361)
Firm Size: 10 - 49	0.0801***	(0.0241)	0.0796***	(0.0243)	0.0854***	(0.0243)
Firm Size: 50 - 199	0.119***	(0.0254)	0.116***	(0.0257)	0.128***	(0.0261)
Firm Size: 200+	0.141***	(0.0276)	0.137***	(0.0282)	0.150***	(0.0286)
Occ: Professionals	-0.0309	(0.0751)	-0.0299	(0.0758)	-0.0433	(0.0760)
Occ: Technicians	-0.180**	(0.0735)	-0.179**	(0.0740)	-0.191**	(0.0745)
Occ: Clerks	-0.352***	(0.0731)	-0.352***	(0.0737)	-0.360***	(0.0742)
Occ: Service and Retail	-0.517***	(0.0743)	-0.513***	(0.0748)	-0.521***	(0.0754)
Occ: Skilled Agro and Fish	-0.557***	(0.0754)	-0.554***	(0.0760)	-0.552***	(0.0764)
Occ: Craft and Related	-0.507***	(0.0755)	-0.506***	(0.0761)	-0.508***	(0.0765)
Occ: Plant and Machine Opps.	-0.461***	(0.0743)	-0.462***	(0.0749)	-0.473***	(0.0755)
Occ: Unskilled	-0.569***	(0.0734)	-0.566***	(0.0739)	-0.576***	(0.0745)
Constant	6.334***	(0.0997)	6.284***	(0.105)	6.153***	(0.113)
Observations	3462		3462		3462	
Individuals	1154		1154		1154	
Within R^2	0.0535		0.0560		0.0613	
Overall R^2	0.469		0.471		0.471	
Between R^2	0.573		0.574		0.574	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Stage 2 Results: Medium Skill Occupations, Homogenous Returns on Observables

VARIABLES	GLS - No Correction		GLS - Hierarchical		GLS - Unrestricted	
Union	0.0827***	(0.0283)	0.215**	(0.0887)	0.213**	(0.0883)
$\mathcal{C}_i$			0.172	(0.202)	-0.0350	(0.209)
Union*$\mathcal{C}_i$					0.439**	(0.187)
$\mathcal{C}_{it}$			-0.182**	(0.0726)	-0.235***	(0.0906)
Union*$\mathcal{C}_{it}$					0.102	(0.103)
Gender	0.145***	(0.0375)	0.150***	(0.0374)	0.153***	(0.0374)
Health Limitations	-0.112*	(0.0675)	-0.0984	(0.0684)	-0.0908	(0.0672)
Education: 12 years	0.270***	(0.0418)	0.257***	(0.0422)	0.261***	(0.0421)
Education: >12 years	0.593***	(0.0725)	0.589***	(0.0744)	0.596***	(0.0742)
Experience	0.0245***	(0.00537)	0.0239***	(0.00565)	0.0222***	(0.00569)
Experience squared	-0.000230**	(9.46e-05)	-0.000214**	(9.52e-05)	-0.000189**	(9.59e-05)
Private sector worker	0.0278	(0.0456)	0.0567	(0.0468)	0.0538	(0.0469)
Ind: Mining	0.318***	(0.0848)	0.300***	(0.0858)	0.302***	(0.0847)
Ind: Manufacturing	0.0582	(0.0466)	0.0553	(0.0461)	0.0565	(0.0458)
Ind: Utilities	0.00396	(0.101)	-0.0212	(0.0999)	-0.0295	(0.0998)
Ind: Construction	0.185***	(0.0647)	0.208***	(0.0643)	0.210***	(0.0641)
Ind: Wholesale, Retail, and Hotels	-0.0116	(0.0499)	-0.00525	(0.0493)	-0.00267	(0.0490)
Ind: Transport, Storage, and Comms.	-0.0289	(0.0625)	-0.0163	(0.0646)	-0.00372	(0.0642)
Ind: Financial Services	0.165***	(0.0618)	0.161***	(0.0609)	0.163***	(0.0607)
Ind: Personal Services	-0.0296	(0.0584)	-0.0461	(0.0578)	-0.0502	(0.0578)
Firm Size: 10 - 49	0.0565*	(0.0325)	0.0444	(0.0324)	0.0406	(0.0327)
Firm Size: 50 - 199	0.154***	(0.0361)	0.123***	(0.0382)	0.117***	(0.0388)
Firm Size: 200+	0.142***	(0.0366)	0.0898**	(0.0415)	0.0797*	(0.0419)
Occ: Clerks	-0.182***	(0.0455)	-0.171***	(0.0454)	-0.167***	(0.0451)
Occ: Service and Retail	-0.292***	(0.0519)	-0.291***	(0.0515)	-0.289***	(0.0514)
Occ: Skilled Agro and Fish	-0.394***	(0.0676)	-0.391***	(0.0687)	-0.393***	(0.0685)
Occ: Craft and Related	-0.380***	(0.0589)	-0.384***	(0.0603)	-0.387***	(0.0602)
Occ: Plant and Machine Opps.	-0.328***	(0.0559)	-0.333***	(0.0558)	-0.334***	(0.0556)
Constant	6.337***	(0.115)	6.321***	(0.115)	6.314***	(0.114)
Observations	1905		1905		1905	
Individuals	635		635		635	
Within R^2	0.0600		0.0749		0.0813	
Overall R^2	0.244		0.249		0.249	
Between R^2	0.321		0.322		0.319	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Stage 2 Results: Both Samples, Heterogeneous Returns on Observables

	All Occupations				Medium Skilled Occupations			
VARIABLES	Union		Non-Union		Union		Non-Union	
Gender	0.204***	(0.0538)	0.142***	(0.0336)	0.217***	(0.0634)	0.130***	(0.0432)
Health Limitations	0.0148	(0.0756)	-0.0769	(0.0563)	0.0617	(0.113)	-0.149*	(0.0839)
Education: 12 years	0.274***	(0.0781)	0.298***	(0.0327)	0.176*	(0.0992)	0.248***	(0.0461)
Education: >12 years	0.880***	(0.105)	0.839***	(0.0697)	0.537***	(0.139)	0.543***	(0.0844)
Experience	0.0432***	(0.0119)	0.0211***	(0.00433)	0.0256**	(0.0111)	0.0201***	(0.00689)
Experience squared	-0.000495***	(0.000163)	-0.000115*	(6.65e-05)	-0.000189	(0.000191)	-0.000162	(0.000114)
Private sector worker	0.0458	(0.0526)	0.00527	(0.0440)	0.0674	(0.0674)	-0.00182	(0.0620)
Ind: Mining	0.306**	(0.126)	0.301***	(0.0809)	0.320**	(0.136)	0.233**	(0.100)
Ind: Manufacturing	0.0397	(0.0768)	0.113***	(0.0339)	-0.0795	(0.0827)	0.0985*	(0.0542)
Ind: Utilities	0.0333	(0.172)	0.195**	(0.0802)	-0.114	(0.232)	0.0234	(0.103)
Ind: Construction	0.0436	(0.100)	0.233***	(0.0428)	-0.0281	(0.119)	0.276***	(0.0714)
Ind: Wholesale, ...	0.00346	(0.0853)	0.120***	(0.0373)	-0.157	(0.105)	0.0410	(0.0553)
Ind: Transport, ...	-0.0244	(0.123)	0.112**	(0.0505)	-0.198	(0.177)	0.0464	(0.0691)
Ind: Financial Services	0.266**	(0.117)	0.213***	(0.0466)	0.0641	(0.126)	0.204***	(0.0689)
Ind: Personal Services	0.00600	(0.0938)	0.0599	(0.0412)	-0.132	(0.105)	-0.00686	(0.0696)
Firm Size: 10 - 49	0.116	(0.0983)	0.0781***	(0.0251)	0.130	(0.0959)	0.0428	(0.0344)
Firm Size: 50 - 199	0.207**	(0.101)	0.119***	(0.0273)	0.233**	(0.0960)	0.124***	(0.0420)
Firm Size: 200+	0.178*	(0.107)	0.157***	(0.0311)	0.196**	(0.0954)	0.0994**	(0.0470)
Occ: Professionals	-0.134	(0.0900)	-0.00905	(0.0998)				
Occ: Technicians	-0.284***	(0.108)	-0.143	(0.0919)				
Occ: Clerks	-0.348***	(0.108)	-0.392***	(0.0911)	-0.0323	(0.0690)	-0.276***	(0.0589)
Occ: Service and Retail	-0.435***	(0.116)	-0.594***	(0.0914)	-0.0874	(0.0762)	-0.449***	(0.0678)
Occ: Skilled Agro ...	-0.625***	(0.155)	-0.585***	(0.0907)	-0.251	(0.158)	-0.498***	(0.0813)
Occ: Craft and Related	-0.462***	(0.129)	-0.563***	(0.0919)	-0.210**	(0.106)	-0.521***	(0.0765)
Occ: Plant and ...	-0.455***	(0.122)	-0.533***	(0.0911)	-0.243***	(0.0924)	-0.454***	(0.0715)
Occ: Unskilled	-0.653***	(0.128)	-0.616***	(0.0896)				
$\mathcal{C}_i$	0.543	(0.479)	-0.908***	(0.280)	0.679*	(0.353)	0.101	(0.234)
$\mathcal{C}_{it}$	0.246	(0.290)	-0.113*	(0.0630)	-0.0847	(0.119)	-0.187*	(0.0999)
Constant	6.164***	(0.328)	6.302***	(0.128)	6.291***	(0.257)	6.526***	(0.128)
Observations	862		2600		513		1392	
Individuals	452		1016		258		543	
Within R^2	0.0518		0.0687		0.0688		0.0717	
Overall R^2	0.443		0.471		0.231		0.256	
Between R^2	0.495		0.551		0.277		0.326	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Union Wage Effects by Estimation Procedure

	Estimation Method	Union Parameter	Union Effect	$\sigma_{0,\alpha\theta}$	$\sigma_{1,\alpha\theta}$	$\sigma_{0,\epsilon\eta}$	$\sigma_{1,\epsilon\eta}$
All Occupations	GLS	0,079***	8,2%				
	GLS – Hierarchical	0,134***	14,3%	-0,426**	-0,426**	-0,0843*	-0,0843*
	GLS – Unrestricted	0,193***	21,3%	-1,010***	1,099***	-0,0675	0,198
	Heterogeneous	0,189***	23,5%	-0.908***	0,543	-0.113*	0,246
Medium Skill Occupations	GLS	0,083***	8,6%				
	GLS – Hierarchical	0,215**	24,0%	0,172	0,172	-0.182**	-0,182**
	GLS – Unrestricted	0,213**	23,7%	-0,035	0,439**	-0,235***	0,102
	Heterogeneous	0,115***	14,4%	0,101	0,679*	-0,187*	-0,085

*** p<0.01, ** p<0.05, * p<0.1

Table 8: Summary of Treatment Effects

		Treatment Parameter	As Percentage
All Occupations	ATE	0,19	23,5%
	TT	0,19	23,0%
	TUT	0,19	23,6%
Medium Skill Occupations	ATE	0,11	14,4%
	TT	0,12	14,7%
	TUT	0,11	14,2%

Figure 3: Union Wage Effect vs Observed Wage Percentile