Project 2

Linear Panel Data and Production Technology

Econometrics B

Mikkel L. Dollerup ndt926

Emil V. Ramsbæk tdh424

Department of Economics

University of Copenhagen

January 9, 2024

Introduction

In this report, we will examine a part of a well-known problem which is union wage premium. Since union membership isn't random, we want to find investigate if there occurs true state dependence in union participation. The report is based on data collected over a 5-year(1980-1984) period covering 545 males in the 18-34 age span. All men worked all years.
Our object of interest is to find out what effect a lagged variable has on the current variable. Furthermore we want to find the best dynamic model for estimating the APE of lagged y (state dependence).
have the model:

$$P(union_{it=1}|union_{i,t-1},...,union_{i0},\mathbf{z}_i,c_i) = G(\mathbf{z}_i t\gamma + \rho \cdot union_{it-1} + c_i)$$

where $t = 1, 2, 3, 4, 5$, and $\mathbf{z}_i$ and c_i are vectors of individual and time-invariant unobserved characteristics, respectively.
Our results indicate that there is a significant level of state dependence in union membership, with individuals who were previously members being more likely to remain members in the future. The magnitude of this effect varies depending on the model specification, but is consistently present across different specifications. These findings have important implications for policymakers and labor unions in understanding the dynamics of union membership and the factors that contribute to its persistence.

Economic theory

When estimating our dynamic model we assume that the lagged variable $union_{it-1}$ accounts for all previous periods. By definition, the explanitory variables z_{it} are strict exogenous.
True state dependence occurs if $\rho \neq 0$ and if $\rho > 0$ the individual has a higher change to be part of a union next period if they were part of one in the previous period.
True state dependence might come from unobserved heterogeneity which is why we want to test for heterogeneity from the unobserved variable, c_i.

$$\prod_{t=1}^{T} G(\mathbf{z_t}\delta + \rho y_{t-1} + c)_i^y (1 - G(\mathbf{z_t}\delta + \rho y_{t-1} + c)^{1-y_i} \tag{1}$$

Since we wish to control for unobserved effects, we wish to use the random effects probit model(RE) which includes the individual specific random effects c_i. Since these unobserved effects are existing we use the dynamic model, that yields the initial conditions problem since $union_{i0}$ is being removed. A correlation between c_i and $union_{i0}$ is likely to occur. To deal with this problem we use the Wooldridge/Champerlain approach. We treat $unio_{i0}$ as a non-random and make a strong assumption and assuming independence between $unio_{i0}$ and c_i. We assume distribution:

$$c|union_0; z \sim N(\psi + \xi_0 y_{i0} + z_i \xi, \sigma_a) \tag{2}$$

Yields to new variable form:

$$y_{it} = 1[z_{it}\delta + \rho y_{it-1} + c_i + e_{it}] = 1[z_{it}\delta + \rho y_{it-1} + \psi + \xi_0 y_{i0} + z_i \xi + a_i + e_{it}]$$

where $a_i \sim N(0, \sigma_a^2)$ and $e_i \sim N(0, 1)$
With these assumptions we use the random effects probit model as:

$$f(y_1, \dots, y_T \mid \mathbf{x_i}; \theta) = \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T} f(y_t \mid \mathbf{x_{it}}, c, \beta_i) \right] \frac{1}{\sigma_a} \phi(\frac{c}{\sigma_a}) dc$$

Link function

$$G(z_{it}\gamma + \rho \cdot union_{it-1} + c_i) = \Phi(z_{it}\gamma + \rho \cdot union_{it-1} + c_i) \tag{3}$$

The probit link function used in this model is defined as the cumulative distribution function of the standard normal distribution, denoted by $\Phi(\cdot)$. The standard normal distribution has a mean of 0 and a variance of 1, and its cumulative distribution function gives the probability of observing a value less than or equal to a given threshold.

The probit link function assumes that the conditional distribution of the dependent variable given the independent variables is a standard normal distribution, implying that the residuals from the model are normally distributed. This assumption is necessary for the maximum likelihood estimation of the model parameters to be valid. Moreover, the linearity assumption in the link function implies that the effect of the independent variables on the dependent variable is constant across the range of values.

Hypothesis testing

We are testing to see if ρ is different to 0, which can tell us if it has an effect being a union member in the last period on the current period.

Null Hypothesis: There is no significant relationship between union membership in the previous period and union membership in the current period.(No state dependence)

$$H_0 : \rho = 0 \tag{4}$$

Alternative Hypothesis: There is a significant relationship between union membership in the previous period and union membership in the current period. (State dependence)

$$H_A : \rho \neq 0 \tag{5}$$

To test if union membership in the previous period has an impact on union membership in the next period, we can use a likelihood ratio test.
Firstly, we estimate the dynamic model with the current and lagged union membership as regressors

$$G(z_{it}\gamma + \rho \cdot union_{it-1} + c_i) = \Phi(z_{it}\gamma + \rho \cdot union_{it-1} + c_i) \tag{6}$$

Then, estimate the restricted model where the lagged union membership is not included:

$$G(z_{it}\gamma + c_i) = \Phi(z_{it}\gamma + c_i) \tag{7}$$

We compute the likelihood of both models and take the ratio of the likelihoods. The test statistic is distributed as χ^2 with 1 degree of freedom under the null hypothesis that the lagged union membership has no impact on the current union membership.

Assuming a significance level of 0.05, and using a chi-squared distribution table with 1 degree of freedom, we find that the critical value for rejecting the null hypothesis is 3.84.

Empirical Analysis

To estimate the level of state dependence in union membership, we begin by selecting an appropriate model that can account for the time dependence in the data. We consider several candidate models.
Looking at linear probability model POLS and FE in Table 1 in Appendix, we see a estimated effect of being married of 0.471 and 0.0012. Looking at our static pooled probit Table 2 APE of 0.04128, we could interpret the FE method heavily underestimates the effect in union participation when being married and a slightly underestimation by POLS. This might be caused by the FE method accounts for fixed unobserved effect over time when marriage changes over time.

Proceeding to the dynamic models. We estimated both the pooled probit and the RE probit model, but as the RE probit model accounts for unobserved heterogeneity, we proceed with that model. The RE probit model in Table 5 include $union_{t-1}$ and c_i the approach is used to handle the initial conditions problem formerly described in Economic Theory. By doing this approach we control for unobserved heterogeneity between x_{it} and c_i.

Comparing the static model in table 3 and dynamic model in table 5 we see a significant difference in heterogeneity since larger σ_a in the static model since the dynamic model accounts for state dependence.

From table 5 we estimate a postive significant ρ value of 0.6843 which means being a union member in previous period has a positive impact on being a member in the next period. APE = 0.122 which means theres a state depedence of 12,2 pct which can in interpretted as being a member this period is 12.2 pct. larger when bering a member in the previous periods.

Our result is backed up by an LR test following a χ^2 distribution with 1 degree of freedom. A LR value of 665,384 with following hypothesis:

$$H_0 : \rho = 0 \ , H_1 : \rho \neq 0$$

$665,348 > 3.84$ means we accepts H_1 and conclude theres state dependence.

Discussion

Our data consist only of males aged 18-34 and from 1980-1984. Males, especially young males, are known for being larger risk-lovers than females, which could form a miss leading picture of union participation. As of the time span, certain economy-periods are more volatile than others, which could lead to more people participating in a union for safety or vice versa.

One assumption of the random effects probit model is that the unobserved individual-specific effects (the random effects) are independent of the observed covariates. This assumption may not hold if there are unobserved factors that affect both the probability of union membership and the observed covariates, such as unobserved ability or motivation. In such cases, the random effects may be correlated with the observed covariates, which violates the assumption of independent random effects.

Another assumption of the random effects probit model is that the random effects are normally distributed. This assumption may not hold if there are extreme values or outliers in the data that deviate from the normal distribution, or if the sample size is too small to accurately estimate the distribution of the random effects.

A further assumption of the random effects probit model is that the error terms are independently and identically distributed. This assumption may not hold if there are correlated errors, such as serial correlation over time, measurement error, or omitted variable bias.

Conclusion

From the different estimates, we can conclude that there is true state dependence when accounting for union-participating. We also observe that the state state dependence also comes from the unobserved effects. After all, an individual is more likely to participate in a union if they we're a member last period.

Appendixes

Table 1: POLS and FE

parnames	$\hat{\beta}_{POLS}$	$\hat{\beta}_{FE}$
married	0.0471(0.0291)	0.0012(0.0246)
per2	0.0203(0.0190)	-0.002(0.0178)
per3	0.0243(0.0211)	0.0053(0.0201)
per4	0.0091(0.0219)	-0.0058(0.0203)
per5	0.0121(0.0235)	-0.004(0.0213)
educ	0.0183(0.0017)	-

Table 2: Static Pooled Probit

parnames	$\hat{\theta}$	se	APE
married	0.12990	0.08968	0.04128
per2	-0.01878	0.05613	-0.0597
per3	-0.00620	0.06355	-0.00197
per4	-0.05099	0.06589	-0.01620
per5	-0.04170	0.07068	-0.01325
educ	-0.00852	0.02212	-0.00271
const	-0.59508	0.26237	-0.18912

Table 3: Static Random effects probit

parnames	θ_{hat}	se	APE
married	0.00504	0.12613	0.00077
married_bar	0.29153	0.25105	0.04478
educ	-0.05135	0.04785	-0.00789
per2	-0.02721	0.09515	-0.00418
per3	0.01839	0.09743	0.00282
per4	-0.06324	0.10791	-0.00971
per5	-0.02262	0.11241	-0.00347
const	-0.86331	0.57549	-0.13261
σ_a	1.82459	0.10898	0.28026

$$LLH_{dynamicmodel} = -796.76$$
$$LLH_{staticmodel} = -1129,43$$

$$LR = 2(-796.76 - (-1129,43)) = 665,34$$

χ^2 with 1 degree of freedom = 3,84

Table 4: Random effects probit with control for x_{it}

parnames	**theta_hat**	**se**	**t-values**	**jac**	**APE**
l1.union	0.68428	0.12835	5.33136	-0.00000	0.10746
married	-0.08670	0.17024	-0.50930	0.00000	-0.01362
educ	-0.01997	0.04251	-0.46989	-0.00000	-0.00314
per3	0.03826	0.10893	0.35119	0.00000	0.00601
per4	-0.05680	0.11614	-0.48905	0.00000	-0.00892
per5	-0.00062	0.11715	-0.00529	-0.00000	-0.00010
const	-1.73793	0.50076	-3.47059	-0.00000	-0.27292
married2	0.15140	0.22894	0.66130	-0.00000	0.02377
married3	-0.03627	0.23044	-0.15738	0.00000	-0.00570
married4	-0.16931	0.13686	-1.23713	-0.00000	-0.02659
married5	0.32654	0.19684	1.65890	0.00000	0.05128
union1	1.80543	0.23198	7.78256	0.00000	0.28352
sigma_a	1.23950	0.13511	9.17422	0.00000	0.19465

Table 5: Dynamic RE probit: Dep. var. = ['union']

parnames	θ_{hat}	se	t-values	jac	APE
l1.union	0.68428	0.12835	5.33136	-0.00000	0.10746
married	-0.08670	0.17024	-0.50930	0.00000	-0.01362
educ	-0.01997	0.04251	-0.46989	-0.00000	-0.00314
per3	0.03826	0.10893	0.35119	0.00000	0.00601
per4	-0.05680	0.11614	-0.48905	0.00000	-0.00892
per5	-0.00062	0.11715	-0.00529	-0.00000	-0.00010
const	-1.73793	0.50076	-3.47059	-0.00000	-0.27292
married2	0.15140	0.22894	0.66130	-0.00000	0.02377
married3	-0.03627	0.23044	-0.15738	0.00000	-0.00570
married4	-0.16931	0.13686	-1.23713	-0.00000	-0.02659
married5	0.32654	0.19684	1.65890	0.00000	0.05128
union1	1.80543	0.23198	7.78256	0.00000	0.28352
$sigma_a$	1.23950	0.13511	9.17422	0.00000	0.19465