Econometrics and Empirical Economics

Empirical Exercise

Lecturer:
Prof. Vishal Kamat

Student:
Jingwan LUO (Student Number 22007276)
Zhihan Zhang (Student Number 21613596)

Contents

Toulouse, December 17, 2021

1 QUESTION 1, LINEAR PROBABILITY MODEL, PROBIT MODEL AND LOGIT MODEL

1.1 3.(a)

From codebook of this dataset, Stock and Watson (2003) gives the following definition, along with our understanding towards them.

- `faminc`: 1988 family income, divided by 1000, unit in $.
- `cigtax`: cigarette tax in home state, 1988, unit in $.
- `cigprice`: cigarette price in home state, 1988, unit in $.
- `bwght`: birth weight, unit in ounces.
- `fatheduc`: father's years of education, unit in years.
- `motheduc`: mother's years of education, unit in years.
- `parity`: birth order of child, variable serves as an index, can take values from 1 to 6, sample average approximately 1.63.
- `male`: gender binary variable, $= 1$ if male child.
- `white`: race binary variable, $= 1$ if race is white.
- `cigs`: cigarettes smoked per day while pregnant, unit of cigarettes.
- `bwghtlbs`: birth weight, unit in pounds.
- `packs`: packs smoked per day while pregnant, unit in packs (if half pack of cigarettes were smoked, put 0.5, i.e. $20 * 0.5 = 10$ cigarettes were smoked).
- `lfaminc`: log of `faminc`.

Possible research angle: Analyze if and how the birth weight depends on `cigs`, `parity`, `faminc`, `motheduc` and `fatheduc`. The original source of the dataset is from Mullahy (1997), which uses IV approach to study the smoking behaviour.

1.2 3. (b)

Please find Table 1 in TABLE ANNEX to be the summary statistics of all variables.

1.3 3. (c)

There are 212 smoking woman in this dataset.

1.4 3. (d)

There are 165 white smoking woman in this dataset.

1.5 3. (e)

There are 53 smoking woman with family income above sample average.

1.6 3. (f)

There are 53 smoking woman with family income above sample median.

1.7 3. (g)

There are 1 such observation with one missing value among `smoke, motheduc, lfaminc, white`.

1.8 4. (a)

Please find the regression table of Linear Probability Model in Table 2 at TABLE ANNEX.

1.9 4. (b)

If not adjusted for robust standard error, all explanatory variables are significant at 5 % level, if robust *s.e.* is used, `white` is not significant at 5 % anymore.

1.10 4. (c)

If `motheduc` increases by one year, then the probability of smoking decreases by -0.0293.
If race is `white`, then the probability of smoking increases by 0.0484.
If `lfaminc` increases by 100 %, then the probability is negatively changed by 0.042. Since:

$$y = \beta_1 \log x_1$$
$$\frac{\partial y}{\partial x_1} \longrightarrow \frac{\beta_1}{100} \approx \frac{\Delta y}{100 \cdot \frac{\Delta x_1}{x_1}}$$

1.11 4. (d)

The proportion of fitted probabilities outside the unit interval is 64/1387 = 0.04614275.

1.12 5. (a)

Please find the result of logit regression in 3.

1.13 5. (b)

Please find the summary of marginal effect on logit model in 4.

1.14 5. (c)

In terms of sign and magnitude, the AME and MEM of logit model are consistent with the result of LPM.

1.15 5. (d)

For each unit in the sample we compute the fitted probability and set the predicted value of `smokes` equal to 1 if the fitted probability is ≥ 0.5 and 0, *i.e.* we define it's a correct prediction if the fitted probability is greater than 0.5. Then, we compute the proportion of correction predictions who did not smoke as:

$$\frac{\sum_{i=1}^{n} \mathbb{1}\{y_i = 0 \text{ and } \hat{y}_i = 0\}}{\sum_{i=1}^{n} \mathbb{1}\{y_i = 0\}} \tag{1.15.1}$$

and the proportion of correct predictions who did smoke as:

$$\frac{\sum_{i=1}^{n} \mathbb{1}\{y_i = 1 \text{ and } \hat{y}_i = 1\}}{\sum_{i=1}^{n} \mathbb{1}\{y_i = 1\}} \tag{1.15.2}$$

For LPM model who did not smoke: $\frac{1174}{1176}$;
who did smoke: $\frac{0}{212}$.
For Logit model who did not smoke: $\frac{1166}{1176}$;
who did smoke: $\frac{3}{212}$.
The two models in use have approximately the same proportions of correct predictions in the current definition of a "correct prediction". Logit model out-performs LPM in the subsample where $smokes = 1$ while LPM out-performs in the subsample where $smokes = 0$.

1.16 5. (e)

In our logit regression model result, we can conclude that, the covariates `motheduc` and `lfaminc` have significant negative effects on `smokes` at least at the 0.1% level of significance, which means that a woman smokes has less probability of smoking during pregnancy as she has a higher level of education, or as a higher family income. However in this non-linear model we cannot interpret the coefficient as constant marginal effects of `motheduc` and `lfaminc`.

In one word, a woman's level of education one year higher will has 3% points smaller probability of smoking during pregnancy than one with a lower education level. And one higher log of family salary will let her has 3.6% smaller probability of smoking. but the woman's skin color is not a significant factor.In order to confirm the credibility of the results shown by our logit model, we reviewed two earlier studies and one recent study on the determinants of smoking during pregnancy. The studies we refer to here are A. Mcknight (1986), P. Nafstad (1996) and Ergin (2010).

A. Mcknight (1986) shows that compared with non-smokers during pregnancy, smokers have lower educational qualifications, are less likely to be employed, and that smokers' partners are more likely to smoke than non-smokers. The study also showed that although close to 65% of smokers knew that smoking would have adverse effects on babies, only 11.5% of smokers actually quit smoking during pregnancy. P. Nafstad (1996) showed that the main factors for smoking during pregnancy are young age, cohabitation with smokers, low education and not breastfeeding. Women with higher education and living with non-smokers had seven times the rate of smoking cessation during pregnancy than women with lower education and living with smokers. Ergin (2010) showed that the mother's young age, low education level, low social class, migrant population and spouse smoking during pregnancy are significant determinants of the mother's higher smoking rate during pregnancy.

Compared with these studies, the influence direction of our education level factor is consistent with them, and it has a significant negative effect on smoking during pregnancy. In Ergin (2010), illiterate women are 3.4 times more likely to smoke during pregnancy than other women. However, we noticed that in these studies, researchers not only used other variables that were different from ours, but also used different models. For example, P. Nafstad (1996) uses multiple logistic regression. The multivariate logistic regression used by Ergin (2010) divides the regression into two logit models, namely "non-smokers" vs. "daily and occasionally smokers", and "daily smokers" vs. "non-smokers, ex-smokers and occasional smokers". This logit regression reflects more details.

We think that there are variables such as breastfeeding, women's age, whether they are aware of the dangers of smoking during pregnancy, social class, and whether their partner smokes or not, which need to be added to the logit regression model. In the above studies, the smoking status of people living with pregnant women also has a significant impact on women's smoking during pregnancy. Therefore, peer effects are a factor that we cannot ignore, and it is related to smoking behavior.

1.17 5. (f)

With `lfaminc` set to sample average, `white` set to be 1, the marginal effect of individual with `motheduc` being 16 is -.0142874, while the marginal effect of individual with `motheduc` being 12 is -.0330342, the estimated difference between the two case is: $(-.0142874-(-.0330342)) = .0187468$.

1.18 5. (g)

The endogeneity issue can be summarised as:

$$\mathbb{E}[\varepsilon|\mathbf{x}] \neq 0$$

The variable `lfaminc` is very likely to be endogenous, since smoking behaviour can be determined by the monthly spending budget of the individual, whereas the `lfaminc` can hardly reflect.
The variable `motheduc` is less likely to be endogenous, since `motheduc` hardly co-depend on other covariates in the system, moreover, as a determined trait of the individual, `motheduc` has a more direct relationship with smoking behaviour.

1.19 5. (h)

We followed the standard two step procedure to test the endogeneity of **lfaminc**:
FIRST STAGE:
Regress **lfaminc** on:

$$motheduc\ fatheduc\ bwght\ male\ white\ lbwght\ cigs\ cigtax\ cigprice$$

and collect the residuals.
SECOND STAGE:
We regress **smokes** on:

$$smokes\ lfaminc\ white\ motheduc\ lfaminc\_res$$

and test for the coefficient on **lfaminc_res**. We have a p value close to zero and reject the null. **Possible solution to such situation**: we could instead choose method of IVprobit. Available instruments for **lfaminc** could be **motheduc** or **fatheduc**.

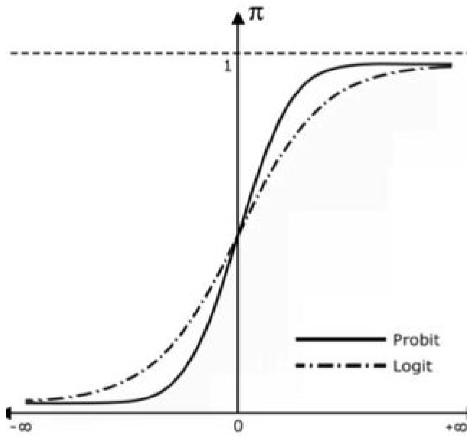
1.20 (6)

Please find the summary of marginal effects on Probit Model in Table 5.
Recall from the formula of Marginal effect (discrete version):

$$\mathbb{P}[Y_i = 1|x_{i[-k]}, x_{ik} + 1] - \mathbb{P}[Y_i = 1|x_{i[-k]}, x_{ik}]$$

Therefore the calculation of marginal effect for discrete variable relies on its underlying CDF.
The results of Probit model is **not far** from Logit model in terms of magnitude and sign, potentially because firstly the marginal effects are either averaged or evaluated at their means. **Secondly**, cumulative distribution functions of normal (which is the underlying distribution of Probit) and logistic

Figure 1: LOGISTIC DISTRIBUTION AND NORMAL DISTRIBUTION



distribution (which is the underlying distribution of Logit) **intersects at** $x = 0$, while at other parts of support of x Normal is not far from Logistic, as illustrated from Figure 1 [1]:

2 QUESTION 2, TOBIT MODEL

2.1 3. (a)

From the codebook of Stock and Watson (2003), here is a brief description of the variable in use:

- `hrbens`: hourly fringe benefits, unit in $.
- `exper`: years work experience, unit in years.
- `age`: age, unit in years.
- `educ`: years schooling, unit in years.
- `tenure`: years employed with current employer, unit in years.
- `married`: binary variable, $= 1$ if married.
- `male`: gender binary variable, $= 1$ if gender is male.
- `white`: race binary variable, $= 1$ if race is white.
- `nrtheast`: living location variable, $= 1$ if live in northeast.
- `nrthcen`: living location variable, $= 1$ if in north central.
- `south`: living location variable, $= 1$ if live in south.
- `union`: binary variable indicates whether the individual is in the union, $= 1$ if union member.

Possible research angle: how hourly benefits of an individual affected by their certain traits? (For instance, we can focus on a comparison between genders holding other traits ceteris paribus; or we can look at the regional effect on hourly benefits).

[1] Source:https://www.researchgate.net/figure/Comparison-of-probit-and-logit-models-as-cumulative-distribution-The-general-form-of-the$_{fig}12_347974918[accessed16Dec, 2021]$

2.2 3. (b)

Please find summary statisitcs of the variable in use from Table 6.

2.3 3. (c)

The number of woman is 212.

2.4 3. (d)

The number of married woman tenure larger than sample average is 53.

2.5 3. (e)

The mean of `hrbens` if the individual is male: 1.061646.
The mean of `hrbens` if the individual is female: .638371.

2.6 (4)

From Table 7, column (1) presents OLS regression, column (2) presents **linear probability model with binary dependent variable** indicating whether `hrbens` is **positive**, column (3) is Tobit result, (4) is Tobit with squared term added.
`educ`, `tenure`, `male` and `union` are significant from OLS (linear regression) result.
`educ`, `union` are significant from LPM result. Interpretations to Linear regressions, `hrbens` not censored (then we are facing with a standard OLS instead of a LPM):

- one year raise in working experience raise hourly benefit by 0.003.
- one year older in age brings down hourly benefit enjoyed by the employee by 0.002.
- one more year of schooling raise hourly benefit enojoyed by 0.08.
- one more year being employed with the current employer raise hourly benefit by 0.03.
- being a white married male compare to the case where not raise hourly benefit by 0.09. 0.089 and 0.3 respectively.
- being employed in the location northeast, north central and south of U.S. brings down hourly benefit enjoyed by -0.08, -0.05 and -0.03 respectively, compares to the case where the worker does not work in these regions.
- being a union memeber raise hourly benefit by 0.4.

However, the above interpretations applies to a standard OLS regression. For interpretation on a linear probabiliy model, I generated an indicator variable of `hrbens`, `b_hrbens`.
The interpretations of LPM in this case is:

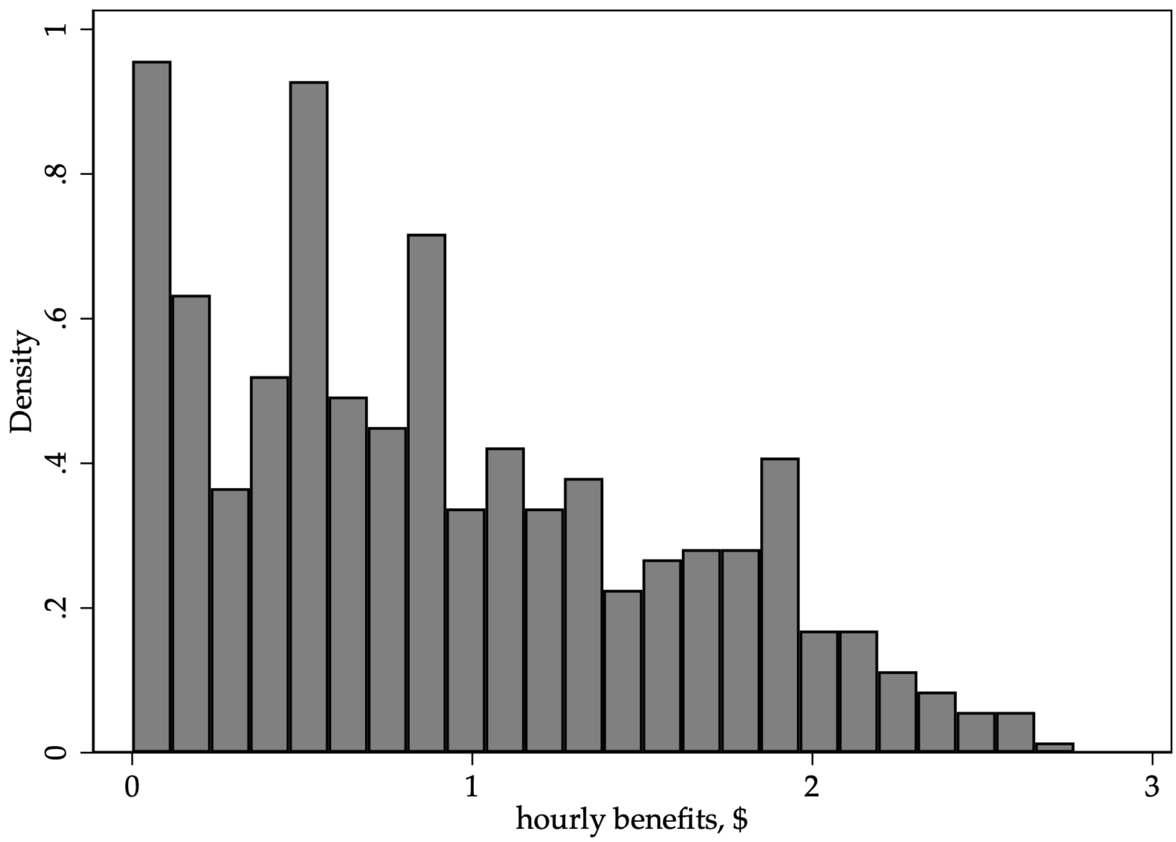
- one year raise in working experience raise the probability of getting a positive hourly benefit is 0.0028.
- one year older in age brings down the probability of enjoying a positive benefit by the employee by 0.0003.
- one more year of schooling raised the predicted probability of positive hourly benefit by 0.01.
- one more year being employed with the current employer raise hourly benefit by 0.0008.

- being a while married male raise the probability of having positive fringe by 0.04, 0.01 and 0.005 respectively.
- being employed in the location northeast, north central and south of U.S. brings changes on the probability of positive fringe by 0.01, -0.004 and 0.004 respectively, compares to the case where the worker does not work in these regions.
- being a union memeber raise probability of getting positive fringe by 0.08.

2.7 (5)

We **can** classify `hrbens` as a censored dependent variable.
Reason: We plot the distribution of `hrbens`, as illustrated in Figure 2. what we could observe is that

Figure 2: Distribution of Hourly Benefit, Whole Sample



there are siginificant amount of observations concentrate on support 0. Therefore the censoring we could make is:

$$y = \max\{y^*, 0\}$$

We **can** classify `hrbens` as a corner solution dependent variable.
Reason: it is a continuous variable except at support 0, moreover the probability mass function is very likely to be strictly positive.

2.8 (6)

Please find the regression result using a tobit model in Table 7, Column (3).

2.9 (7)

From standard theory we know that OLS estimates are generally downward biased than Tobit model:

$$\begin{aligned}\mathbb{E}[\mathbf{y} \mid \mathbf{y}\ observed] &= \mathbb{E}[\mathbf{y} \mid \mathbf{x},\ \mathbf{y} > 0] \\ &= \mathbb{E}[\mathbf{y} \mid \mathbf{x}, \varepsilon > -\mathbf{x}\beta] \\ &= \mathbf{x}\beta - \sigma \left(\frac{\phi\left(-\frac{\mathbf{x}\beta}{\sigma}\right)}{\Phi\left(-\frac{\mathbf{x}\beta}{\sigma}\right)} \right)\end{aligned}$$

And a OLS model, $\tilde{\beta}$ would be estimated by such a regression model:

$$\mathbf{y} = \mathbf{x}\tilde{\beta} + \delta$$

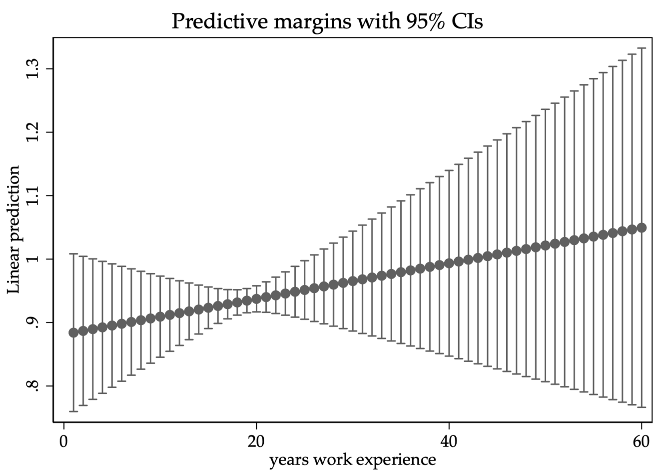
and the composition of the bias is:

$$\mathbb{E}\left[\tilde{\beta}\right] = \beta - \sigma\left(\mathbf{x}'\mathbf{x}\right)^{-1}\mathbb{E}\left[\mathbf{x}'\lambda\left(-\frac{\mathbf{x}\beta}{\sigma}\right)\right]$$

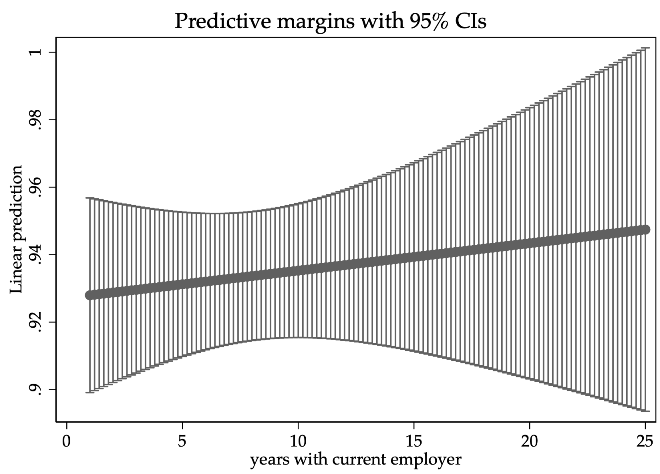
The fact that Tobit and OLS obtain similar results suggest that in our model we may have a close to zero σ of the error term, or a really small inverse Mill's ratio, suggesting the **non-selection hazard** in our model is small.

2.10 (8)

Comparison between the regressions made in Table 7:
In terms of both sign, magnitude or significance, the two model present similar results, except for a comparably larger difference between the variable `tenure` and `exper`.
Conclusion: we should not add these squared terms, since the span of the data does not exhibit a clear non-linear relationship (as there is not enough curvature observed from both fitted plot and margin plot) as depicted in Figure 3 and 4 [2]:



(a) Data Span of Variable *exper*



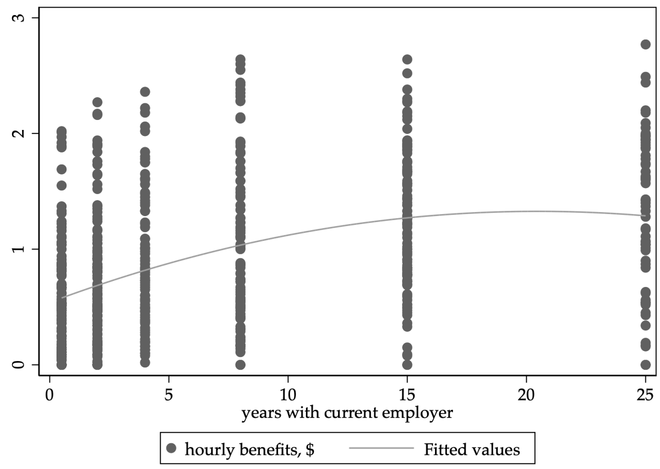
(b) Data Span of Variable *tenure*

Figure 3: DATA SPAN OF SQUARED TERMS

From Figure 4(b) we can see that tenure is better to be censored instead of adding square terms. As for exper, from Figure 4(a) the quadratic relationship is also not clear.

[2] The method follows the suggests of Statalist Discussion Forum: https://www.statalist.org/forums/forum/general-stata-discussion/general/1376936-testing-whether-to-include-a-squared-term

(a) Scatter Plot with quadratic fit of Variable *exper*



(b) Scatter Plot with quadratic fit of Variable *tenure*

Figure 4: SCATTER PLOT OF TARGET VARIABLE'S RELATIONSHIP WITH TENURE, EXPER

2.11 (9)

With average `tenure` squared to be 60.087659 and average `exper` squared to be 346.58822:

$$\mathbb{E}(hrbens|hrbens > 0, \bar{X} = \bar{x}_n) = .0508968$$

2.12 (10)

FIRST STAGE:
We run an OLS regression of `educ` on all the exogenous covariates and we save the residuals.
SECOND STAGE:
We run a Tobit of `hours` on the rest of the covariates except for location binary variables from the original model.
From the second stage we find that the coefficient of residuals saved from first stage is not significant at 5 % or even 10 %, therefore, we cannot reject that `educ` is exogenous.

2.13 (11)

One of the famous literature from labor topic is Vella (1993) which focusing on illustrating the censored regressors issue using fringe data as in their example A. From their linear regression result, the coefficient on male is approximately 0.3 (0.287, significant at 10 % level specifically in their Table 2, Column 1, reduced-form OLS). This result is very similar to what we have obtained for coefficient of 0.252, significant at 1 % level result. Moreover, from the Vella (1993), the endogeity issue of `fringe` is highlighted since it is a variable that is simultaneously determined by wage. The fringe in the paper was censored at 0, and by the data span of our variable `hrbens`, it can also be viewed as censored since the variable is 0 as depicted in Figure 2 previously, that also explains why the regression result on regressor `male` is so similar. Alternatively, the endogenous relationship of fringe and wage could be another possible aspect that we could look at,

$$\log \text{fringe} = \alpha + \sum \alpha_j \cdot \text{personal characters} + \\ \sum \alpha_l \cdot \text{regional dummies} + \sum \alpha_f \cdot \text{industry dummies} + \alpha \cdot \text{fringe}$$

to deal with the endogeneity issue here in this exercise, we could follow the approach listed in the paper, and to estimate the reduced form OLS assuming the error is normal:

$$\text{fringe} = \beta + \sum \beta_j \cdot \text{personal characters} + \\ \sum \beta_l \cdot \text{regional dummies} + \sum \beta_f \cdot \text{industry dummies}$$

and then estimate β by Tobit. The authors also tested the endogenous regressor issue and concluded that the optimal approach to test is to use a LM test.
Another famous paper by Solberg and Laughlin (1995) also concluded that gender issue is a significant determinant for the distinctions found in hourly benefit, this coincides with our findings in the exercise, as coefficient of `male` is greatly larger than the others and obtain significance even at 1 % level. Finally, interesting paper by Rand and Tarp (2011) discusses the gender difference on fringe in the context of Vietnamese SME employers, uses a two-stage model on the elabration of this probit baseline:

$$b_i = \alpha_0 + x_i\alpha_1 + \bar{w}_i\alpha_2 + \bar{s}_i\alpha_3 + e_i$$

where b_i is a binary variable of positive fringe, $\bar{w}_i$ is the average wage, x_i is benefit proxy and $\bar{s}_i$ is skill level of the worker, and the result of their coefficients on gender is all positive and significant at 1 % level, suggesting gender is indeed an important factor to affect fringe, and female employer are more likely to provide fringe benefits such as maternity leave for their employees than male workers.

References

A. Mcknight, J.D. Merret (1986). "Smoking in pregnancy a health education problem". In: *Belfast Cohort Study (Northern Ireland)*.

Vella, Francis (1993). "A simple estimator for simultaneous models with censored endogenous regressors". In: *International Economic Review*, pp. 441–457.

Solberg, Eric and Teresa Laughlin (1995). "The gender pay gap, fringe benefits, and occupational crowding". In: *ILR Review* 48(4), pp. 692–708.

P. Nafstad, J.J.K. Jaakkola et al. (1996). "Breastfeeding, maternal smoking and lower respiratory tract infections". In: *ERS Journals Ltd*.

Mullahy, John (1997). "Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior". In: *Review of Economics and Statistics* 79(4), pp. 586–593.

Stock, H and W Watson (2003). "Instructional Statadatasets for econometrics". In: *Boston College Department of Economics*.

Ergin I., Hassoy et al. (2010). "Maternal age, education level and migration: Socioeconomic determinants for smoking during pregnancy in a field study from Turkey". In: *BMC Public Health 10, 325*.

Rand, John and Finn Tarp (2011). "Does gender influence the provision of fringe benefits? Evidence from Vietnamese SMEs". In: *Feminist economics* 17(1), pp. 59–87.

3 TABLE ANNEX

Table 1: SUMMARY STATISTICS FOR EXERCISE 1 DATASET, BWGHT

	Mean	**s.d.**	**min**	**max**	**Observations**
faminc	29.02666	18.73928	.5	65	1388
cigtax	19.55295	7.795598	2	38	1388
cigprice	130.559	10.24448	103.8	152.5	1388
bwght	118.6996	20.35396	23	271	1388
fatheduc	13.18624	2.745985	1	18	1192
motheduc	12.93583	2.376728	2	18	1387
parity	1.632565	.8940273	1	6	1388
male	.5208934	.4997433	0	1	1388
white	.7845821	.4112601	0	1	1388
cigs	2.087176	5.972688	0	50	1388
lbwght	4.760031	.1906622	3.135494	5.602119	1388
bwghtlbs	7.418723	1.272123	1.4375	16.9375	1388
packs	.1043588	.2986344	0	2.5	1388
lfaminc	3.071271	.9180645	-.6931472	4.174387	1388

Table 2: LINEAR PROBABILITY MODEL REGRESSION RESULT

	LPM	**LPM** adj.
	smokes	smokes
motheduc	-0.0293***	-0.0293***
	(0.00433)	(0.00417)
white	0.0484*	0.0484
	(0.0244)	(0.0253)
lfaminc	-0.0427***	-0.0427**
	(0.0119)	(0.0133)
_cons	0.626***	0.626***
	(0.0537)	(0.0565)
N	1387	1387
adj. R^2	0.061	0.061

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: LOGIT REGRESSION RESULT

	(1)
	smokes
motheduc	-0.252***
	(0.0372)
$white = 1$	0.344
	(0.200)
lfaminc	-0.296***
	(0.0866)
_cons	2.013***
	(0.447)
N	1387
adj. R^2	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: LOGIT MODEL: MARGINAL EFFECTS SUMMARY

	Average Marginal Effect	**Marginal Effect at Means**
motheduc	-0.030**	-0.029**
$white = 1$	0.039***	0.037***
lfaminc	-0.036	-0.034

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: PROBIT MODEL: MARGINAL EFFECTS SUMMARY

	Average Marginal Effect	**Marginal Effect at Means**
motheduc	-0.031***	-0.031***
$white = 1$	0.039***	0.038***
lfaminc	-0.031***	-0.031***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: SUMMARY STATISTICS FOR EXERCISE 2 DATASET, FRINGE

	Mean	**s.d.**	**min**	**max**	**count**
hrbens	.909789	.6650045	0	2.77	616
exper	18.61688	12.32906	0	60	616
age	37.65422	12.69757	16	80	616
educ	12.51461	2.733076	6	18	616
tenure	7.751623	7.775509	.5	25	616
married	.6866883	.4642169	0	1	616
male	.6412338	.4800282	0	1	616
white	.9058442	.2922827	0	1	616
nrtheast	.1964286	.3976192	0	1	616
nrthcen	.3035714	.4601734	0	1	616
south	.3474026	.4765319	0	1	616
union	.3181818	.466149	0	1	616

Table 7: REGRESSION RESULT OF EXERCISE 2

	OLS	**LPM w. b. d.**	**Tobit**	**Tobit w. sq.**
	hrbens	binary h.	hrbens	hrbens
exper	0.00299	0.00280	0.00406	0.0307***
	(0.00425)	(0.00350)	(0.00466)	(0.00853)
age	-0.00225	-0.000393	-0.00259	-0.00403
	(0.00415)	(0.00348)	(0.00444)	(0.00434)
educ	0.0822***	0.0131***	0.0869***	0.0803***
	(0.00851)	(0.00380)	(0.00882)	(0.00870)
tenure	0.0282***	0.000812	0.0287***	0.0581***
	(0.00371)	(0.00153)	(0.00372)	(0.0105)
married	0.0899	0.0411	0.103	0.0715
	(0.0499)	(0.0278)	(0.0538)	(0.0529)
male	0.252***	0.0111	0.256***	0.256***
	(0.0497)	(0.0268)	(0.0552)	(0.0539)
white	0.0989	0.00559	0.0994	0.0907
	(0.0721)	(0.0372)	(0.0786)	(0.0769)
nrtheast	-0.0834	0.0139	-0.0778	-0.0480
	(0.0724)	(0.0307)	(0.0775)	(0.0760)
nrthcen	-0.0493	-0.00401	-0.0489	-0.0337
	(0.0627)	(0.0297)	(0.0714)	(0.0698)
south	-0.0285	0.000461	-0.0247	-0.0175
	(0.0653)	(0.0313)	(0.0709)	(0.0693)
union	0.377***	0.0847***	0.403***	0.387***
	(0.0535)	(0.0163)	(0.0523)	(0.0511)
sexper				-0.000552***
				(0.000149)
stenure				-0.00133**
				(0.000410)
N	616	616	616	616
adj. R^2	0.360	0.060		

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$