

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ
ШКОЛА ЭКОНОМИКИ»**

**ФАКУЛЬТЕТ ЭКОНОМИЧЕСКИХ НАУК
ДЕПАРТАМЕНТ ПРИКЛАДНОЙ ЭКОНОМИКИ**

ДОМАШНЕЕ ЗАДАНИЕ №2

по курсу “Микроэконометрика качественных данных”

Выполнила:
студент группы МЭЭП211
Калинин Кирилл Алексеевич

Преподаватель:
к.э.н., преподаватель
департамента прикладной экономики
факультета экономических наук НИУ ВШЭ
Потанин Богдан Станиславович

Москва
2022

Оглавление

Часть 1. Теория и гипотезы.	3
Часть 2. Модель Тобина.	5
Часть 3. Модель Хекмана	11
Часть 4. Модель Ньюи	17
Приложение.....	18

Часть 1. Теория и гипотезы.

1.1. Выберите независимые переменные для уравнения зарплаты и уравнения занятости. Кратко теоретически обоснуйте выбор каждой из них: не обязательно со ссылками на литературу, достаточно здравого смысла. Укажите и кратко обоснуйте предполагаемые направления эффектов. Уравнение занятости должно включать по крайней мере одну переменную, которой не было в уравнении зарплаты, и одну переменную, которая есть в уравнении зарплаты. Желательно, чтобы общая для двух уравнений переменная была непрерывной, например, возраст, а также не использовать более трех переменных в каждом из уравнений.

Для уравнения заработной платы (*wage* - зависимая переменная) были выбраны следующие переменные:

- *age* – возраст женщины (также в уравнение включен квадрат возраста age^2);
- *city* – дамми на проживании женщины в городе;
- *unem* – уровень безработицы в регионе проживания.

Использование выбранных факторов объясняется их предположительным влиянием на уровень заработной платы женщины. Так, предполагается, что зарплата имеет нелинейную зависимость от возраста. В начале карьерного пути при увеличении возраста заработная плата будет расти, так как женщина набирается опыта, становится более компетентной в своей области, а следовательно, больше ценится как специалист. Однако при достижении определенного порогового возраста заработная плата с каждым годом начинает снижаться. Это происходит по причине того, что женщина уже близится к категории пожилых, ее продуктивность падает из-за возможной медлительности и потери былой энергичности, а такой работник меньше ценится на рынке труда. Иными словами, возможные *направления эффектов* следующие: $\beta_{age} > 0$, $\beta_{age^2} < 0$, где β_{age} и β_{age^2} – коэффициенты при переменных *age* и age^2 , соответственно.

Факт проживания женщины в городе, вероятно, будет положительно сказываться на величине ее заработной платы. Это происходит по причине различий в оплате труда в зависимости от места проживания и от характера труда. Зачастую в городах располагаются крупные фирмы и предприятия, туда стекаются образованные работники, которые получают более высокие зарплаты, при этом в областях нет такого спроса на квалифицированную рабочую силу, заработные платы там ниже. Таким образом, предположительное *направление эффекта* положительное, то есть $\beta_{city} > 0$, где β_{city} – коэффициент при дамми переменной *city*.

Уровень безработицы в округе проживания, предположительно, также будет оказывать отрицательное влияние на уровень заработной платы женщины. Этот показатель, скорее всего, будет отражать общую экономическую ситуацию в округе нахождения: чем хуже экономическое положение, тем выше уровень безработицы, меньше готовность предлагать высокие заработные платы работникам. То есть возможное *направление эффекта отрицательное*, иными словами, $\beta_{unem} < 0$, где β_{unem} – коэффициент при переменной *unem*.

Для уравнения занятости (*lfp* - зависимая переменная) были выбраны следующие переменные:

- ***huswage*** – заработная плата мужа женщины;
- ***mrt*** – налоговая нагрузка на женщину;
- ***unem*** – уровень безработицы в округе проживания.

Использование выбранных факторов объясняется их предположительным влиянием на вероятность занятости женщины. Так, предположительно, высокая заработная плата будет снижать вероятность женщины работать. Когда доход мужа очень низкий, вполне вероятно, что женщина пойдет работать, чтобы помочь мужчине обеспечить семью. Далее, с ростом заработной платы мужа, необходимость в том, чтобы женщина работала, падает, так как у мужчины появляется достаточный объем средств для обеспечения жены и детей, и поэтому женщина может оставаться дома и присматривать за детьми. Таким образом, возможное *направление эффекта отрицательное*, то есть $\beta_{huswage} < 0$, где $\beta_{huswage}$ – коэффициент при переменной *huswage*.

Величина налоговой ставки, скорее всего, будет также иметь отрицательное влияние на вероятность занятости. Чем выше налоговая ставка, тем *большую* часть дохода приходится выплачивать в качестве налога, что снижает мотивацию к работе. Иными словами, предположительное *направление эффекта отрицательное*, то есть $\beta_{mrt} < 0$, где β_{mrt} – коэффициент при переменной *mrt*.

Уровень безработицы в округе проживания, вероятно, негативно будет сказываться на вероятности занятости женщины. Так, высокий уровень безработицы свидетельствует о недостатке рабочих мест для людей. Кроме того, при плохой экономической ситуации в месте проживания и, как следствие, высоком уровне безработицы, скорее всего, будут предлагаться низкие заработные платы, которые не будут окупать затрачиваемых усилий и снизят желание женщины работать. Таким образом, вероятное *направление эффекта отрицательное*, то есть $\beta_{unem} < 0$, где β_{unem} – коэффициент при переменной *unem*.

Часть 2. Модель Тобина.

2.1. Оцените Тобит модель, предварительно записав максимизируемую функцию правдоподобия. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

В модели Тобина предполагается следующий вид зависимой переменной (зарплаты):

$$WAGE_i = \begin{cases} WAGE_i^*, & \text{если } WAGE_i^* > 0 \\ 0, & \text{если } WAGE_i^* \leq 0 \end{cases},$$

где $WAGE_i$ – наблюдаемая заработная плата, которую в действительности получает индивид, который работает;

$WAGE_i^*$ – теоретическая, или реальная, заработная плата, которую получал бы индивид, если бы работал.

$WAGE_i^*$ - латентная переменная, которая задается уравнением:

$$WAGE_i^* = \beta_{intercept} + \beta_{age} * age_i + \beta_{age2} * age_i^2 + \beta_{city} * city_i + \beta_{unem} * unem_i + \varepsilon_i,$$

где $\beta_{intercept}$ – константа, или коэффициент при столбце из единиц;

$\varepsilon_i \sim N(0; \sigma^2)$, случайные ошибки, независимы;

$i = 1, \dots, n$, где n – количество наблюдений.

Соответственно, данная модель строится на основе цензурированной выборки, где представлена частичная, или цензурированная, информация, так как заработная плата наблюдаема лишь для тех индивидов, которые работают.

Для оценки модели используется метод максимального правдоподобия и максимизируется следующая функция:

$$L = \prod_{WAGE_i > 0} \varphi_{WAGE_i^*}(wage_i) * \prod_{WAGE_i = 0} P(WAGE_i^* \leq 0) \rightarrow \max(\hat{\beta}, \hat{\sigma}),$$

где $\varphi(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$, функция плотности нормальной случайной величины;

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_{intercept} \\ \hat{\beta}_{age} \\ \hat{\beta}_{age2} \\ \hat{\beta}_{city} \\ \hat{\beta}_{unem} \end{pmatrix}, \text{ вектор оценок коэффициентов, получаемых при максимизации}$$

функции правдоподобия;

$\hat{\sigma}$ - оценка стандартного отклонения случайной ошибки.

Результаты оценивания Тобит модели приведены в *Таблице 1*:

```
Call:
crch(formula = wage ~ age + I(age^2) + city + unem, data = dd, left = 0, right =
Inf)

Standardized residuals:
      Min       1Q   Median       3Q      Max
-0.3051  -0.1460   0.1682   0.5959   4.8032

Coefficients (location model):
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -8.810828   5.745431  -1.534   0.1251
age          0.499116   0.272775   1.830   0.0673 .
I(age^2)     -0.006270   0.003155  -1.987   0.0469 *
city         0.560225   0.420152   1.333   0.1824
unempl       -0.031619   0.064975  -0.487   0.6265

Coefficients (scale model with log link):
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.60826   0.03743   42.97 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Distribution: gaussian
Log-likelihood: -1539 on 6 Df
Number of iterations in BFGS optimization: 37
```

Таблица 1. Результаты оценивания Тобит модели.

Полученные результаты оценивания Тобит модели, представленные в *Таблице 1*, будут проинтерпретированы ниже, в пункте 2.3.

2.2. Опишите преимущества Тобит модели над моделью с усечённой регрессией. Объясните, в каких случаях можно использовать усеченную регрессию, но не получится использовать Тобит модель: приведите гипотетический (можно использовать фантазию) пример.

Преимуществом Тобит модели над моделью с усеченной регрессией является более высокая эффективность оценок. Это происходит по причине того, что усечение влечет за собой большую потерю информации, чем цензурирование, поскольку в первом случае все данные за порогом теряются. При усечении часть значений зависимой переменной не наблюдается совсем, а при цензурировании в данных случаях пропуски заполняются тем показателем, который является порогом усечения в модели с усеченной регрессией и порогом цензурирования в Тобит модели.

Ситуациями, когда можно использовать усеченную регрессию, но не получится использовать Тобит модель, могут быть случаи, когда в действительности часть наблюдений недоступна (не наблюдаема) и была бы неоправданной замена данных пропущенных значений границей цензурирования/усечения.

Так, например, в случае проведения анкетирования людей в интернете, где главным вопросом является количество часов, проводимых в сети интернет, часть наблюдений будет отсутствовать совсем, так как люди, не пользующиеся интернетом, не смогут пройти данный опрос. В этом случае мы не сможем заменить ненаблюдаемые переменные на нули и оценить Тобит модель, так как объясняющие переменные у нас также будут отсутствовать, то есть данных для оценки Тобит модели будет недостаточно. Однако в этом случае подойдет модель с усечённой регрессией, где порогом усечения выступает ноль.

2.3. Проинтерпретируйте полученные значения оценок для каждой независимой переменной. Поясните, как полученные результаты соотносятся с высказанными вами ранее предположениями.

Согласно результатам оценивания модели, приведенным в *Таблице 1*, значимо отличным от нуля оказался возрастной фактор, причем как его линейная часть (на уровне значимости 0.1), так и квадрат возраста (на уровне значимости 0.05). При этом можно обратить внимание на тот факт, что знаки при данных коэффициентах совпадают с предполагаемыми нами изначально. Так, наблюдаемая заработная плата, при прочих равных, скорее всего, имеет квадратичную зависимость от возраста: сначала при увеличении возраста наблюдаемая заработная плата растет, а после определенного возраста эффект от увеличения возраста обратный. При этом для теоретической заработной платы можно интерпретировать значение данных коэффициентов. Так, теоретическая заработная плата, при прочих равных, сначала при увеличении возраста растет, а после 39.8 лет эффект от увеличения возраста обратный. Другие коэффициенты и эффект от относящихся к ним переменных не поддаются интерпретации, так как являются незначимыми, что не позволяет нам на основе данной модели говорить о наличии статистических свидетельств в пользу того, что факт проживания в городе или уровень безработицы в округе влияют на заработную плату женщины.

2.4. Для индивида с произвольными характеристиками укажите (предварительно записав используемые для расчетов формулы):

A) $E(y^*)$

$$E(WAGE_i^*) = \beta_{intercept} + \beta_{age} * age_i + \beta_{age2} * age_i^2 + \beta_{city} * city_i + \beta_{unem} * unem_i.$$

Пусть индивид с произвольными характеристиками имеет следующие значения параметров, представленные в *Таблице 2* в *Приложении*: женщина возраста 35 лет ($age = 35$), проживающая в городе ($city = 0$), который находится в округе с уровнем безработицы 9% ($unempl = 9$).

Для такого индивида математическое ожидание теоретической заработной платы равно 1.252835 доллара/час.

Б) $E(y)$

$$\begin{aligned} E(WAGE_i) &= E(WAGE_i | WAGE_i^* > 0) * P(WAGE_i^* > 0) + E(WAGE_i | WAGE_i^* \leq 0) * \\ &* P(WAGE_i^* \leq 0) = (x_i^T \beta + \sigma \lambda_i) * \Phi\left(\frac{x_i^T \beta}{\sigma}\right) + 0 * \Phi\left(-\frac{x_i^T \beta}{\sigma}\right) = \\ &= (x_i^T \beta + \sigma \lambda_i) * \Phi\left(\frac{x_i^T \beta}{\sigma}\right), \end{aligned}$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$, функция распределения стандартной нормальной случайной величины;

$$x_i = \begin{pmatrix} 1 \\ age_i \\ age_i^2 \\ city_i \\ unem_i \end{pmatrix}, \text{ вектор переменных; } \beta = \begin{pmatrix} \beta_{intercept} \\ \beta_{age} \\ \beta_{age2} \\ \beta_{city} \\ \beta_{unem} \end{pmatrix}, \text{ вектор коэффициентов;}$$

$$\lambda_i = -\frac{\varphi\left(-\frac{x_i^T \beta}{\sigma}\right)}{1 - \Phi\left(-\frac{x_i^T \beta}{\sigma}\right)}; \quad \sigma - \text{стандартное отклонение случайной ошибки.}$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, математическое ожидание наблюдаемой заработной платы равно 2.681149 доллара/час.

В) Вероятность того, что индивид работает

$$P(WAGE_i^* > 0) = \Phi\left(\frac{x_i^T \beta}{\sigma}\right).$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, вероятность занятости равна 0.5990395.

2.5. Для индивида с произвольными характеристиками рассчитайте предельный эффект любой переменной (не дамми), входящей линейно (предварительно записав используемые для расчетов формулы) на:

A) $E(y^*)$

Непрерывной переменной, входящей в модель линейно, является переменная *unempl*, предельный эффект которой и будет рассчитан далее.

$$\frac{\partial E(WAGE_i^*)}{\partial unempl} = \beta_{unempl}.$$

Для индивида с характеристиками, представленными в Таблице 2 в Приложении, предельный эффект (если быть точнее, его оценка) уровня безработицы на математическое ожидание теоретической зарплаты равен -0.03161907.

B) $E(y)$

$$\frac{\partial E(WAGE_i)}{\partial unempl} = \beta_{unempl} * P(WAGE_i^* > 0) = \beta_{unempl} * \Phi\left(\frac{x_i^T \beta}{\sigma}\right).$$

Для индивида с характеристиками, представленными в Таблице 2 в Приложении, предельный эффект (если быть точнее, его оценка) уровня безработицы на математическое ожидание наблюдаемой зарплаты равен -0.01894108.

B) Вероятность того, что индивид работает

$$\frac{\partial P(WAGE_i^* > 0)}{\partial unempl} = \frac{\beta_{unempl}}{\sigma} * \varphi\left(\frac{x_i^T \beta}{\sigma}\right).$$

Для индивида с характеристиками, представленными в Таблице 2 в Приложении, предельный эффект (если быть точнее, его оценка) уровня безработицы на вероятность занятости равен -0.002447566.

2.6*. Добавьте в модель нелинейный эффект (например, квадрат). Повторите предыдущий пункт для переменной, имеющей нелинейный эффект.

В модели присутствует переменная, входящая нелинейно, - возраст: *age* и *age*². Предельные эффекты возрастного фактора следующие:

A) $E(y^*)$

$$\frac{\partial E(WAGE_i^*)}{\partial age} = \beta_{age} + 2 * age * \beta_{age^2}.$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, предельный эффект (если быть точнее, его оценка) возраста на математическое ожидание теоретической зарплаты равен *0.06019851*.

Б) $E(y)$

$$\begin{aligned}\frac{\partial E(WAGE_i)}{\partial age} &= (\beta_{age} + 2 * age * \beta_{age2}) * P(WAGE_i^* > 0) = \\ &= (\beta_{age} + 2 * age * \beta_{age2}) * \Phi\left(\frac{x_i^T \beta}{\sigma}\right).\end{aligned}$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, предельный эффект (если быть точнее, его оценка) возраста на математическое ожидание наблюдаемой зарплаты равен *0.03606128*.

В) Вероятность того, что индивид работает

$$\frac{\partial P(WAGE_i^* > 0)}{\partial age} = \frac{\beta_{age} + 2 * age * \beta_{age2}}{\sigma} * \varphi\left(\frac{x_i^T \beta}{\sigma}\right).$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, предельный эффект (если быть точнее, его оценка) возраста на вероятность занятости равен *0.004659841*.

Часть 3. Модель Хекмана

3.1. Оцените модель Хекмана с помощью метода максимального правдоподобия, предварительно записав максимизируемую функцию правдоподобия и указав независимые переменные в уравнении занятости, которое должно иметь по крайней мере одну переменную, не входящую в уравнение зарплаты. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

Модель Хекмана, в отличие от модели Тобина, предполагает, что вероятность занятости и наблюдаемая зарплата могут описываться разными процессами. Так, имеем два уравнения системы, задающей латентные переменные заработной платы и вероятности занятости, соответственно:

$$\begin{cases} WAGE_i^* = \beta_{intercept} + \beta_{age} * age_i + \beta_{age2} * age_i^2 + \beta_{city} * city_i + \beta_{unem} * unem_i + \\ \varepsilon_i LFP_i^* = \gamma_{intercept} + \gamma_{huswage} * huswage_i + \gamma_{mrt} * mrt_i + \gamma_{unem} * unem_i + v_i \end{cases},$$

где $\begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \right), \rho = corr(\varepsilon_i, v_i), \forall i \in \{1, n\}.$

Тогда факт занятости определяется как:

$$LFP_i = \begin{cases} 1, \text{ если } LFP_i^* > 0 \\ 0, \text{ если } LFP_i^* \leq 0 \end{cases},$$

а наблюдаемая заработная плата определяется системой:

$$WAGE_i = \begin{cases} WAGE_i^*, \text{ если } LFP_i = 1 \\ \text{не наблюдаем, если } LFP_i = 0 \end{cases}$$

Тогда максимизируемая функция правдоподобия имеет вид:

$$L = \prod_{LFP_i^* > 0} \varphi_{\varepsilon_i} \left(\frac{WAGE_i - x_i^T \beta}{\sigma} \right) * \Phi \left(\frac{\rho(WAGE_i - x_i^T \beta)}{\sigma \sqrt{1 - \rho^2}} \right) * \prod_{LFP_i^* \leq 0} (1 - \Phi(z_i^T \gamma)),$$

где $z_i = \begin{pmatrix} 1 \\ huswage_i \\ mrt_i \\ unem_i \end{pmatrix}$, вектор переменных уравнения занятости;

$$\gamma = \begin{pmatrix} \gamma_{intercept} \\ \gamma_{huswage} \\ \gamma_{mrt} \\ \gamma_{unem} \end{pmatrix}, \text{ вектор коэффициентов уравнения занятости.}$$

Результаты оценивания модели Хекмана представлены в *Таблице 3*:

<p>Tobit 2 model (sample selection model)</p> <p>Maximum Likelihood estimation</p> <p>Newton-Raphson maximisation, 6 iterations</p>

```

Return code 2: successive function values within tolerance limit
Log-Likelihood: -1594.029
753 observations (325 censored and 428 observed)
11 free parameters (df = 742)
Probit selection equation:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.012012	0.715763	8.399	0.000000000000000227 ***
huswage	-0.122376	0.018380	-6.658	0.0000000000053910531 ***
mrt	-7.291093	0.898295	-8.117	0.0000000000000001990 ***
unem	0.003904	0.015425	0.253	0.8

```

Outcome equation:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.86122202	4.74089414	0.814	0.4156
age	0.00735266	0.22591471	0.033	0.9740
I(age^2)	-0.00002239	0.00262351	-0.009	0.9932
city	0.80026451	0.33700867	2.375	0.0178 *
unem	0.01469033	0.05397965	0.272	0.7856

```

Error terms:

```

	Estimate	Std. Error	t value	Pr(> t)
sigma	3.34743	0.12800	26.152	< 0.000000000000000002 ***
rho	-0.27556	0.09815	-2.807	0.00513 **

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Таблица 3. Результаты оценивания модели Хекмана на основе метода максимального правдоподобия.

Согласно результатам оценивания модели Хекмана на основе ММП, представленным в *Таблице 3*, в уравнении занятости значимо отличными от нуля на любом разумном уровне значимости оказались коэффициенты при заработной плате мужа и ставке налога. При этом коэффициенты имеют отрицательное значение, что соотносится с первоначальными предположениями. Так, скорее всего, при прочих равных, рост заработной платы мужа и/или ставки налога будет снижать вероятность женщины выйти на работу. Коэффициент при безработице не поддается интерпретации, так как является незначимым, что не позволяет нам на основе данной модели говорить о наличии статистических свидетельств в пользу того, что уровень безработицы в округе влияют на вероятность женщины быть занятой.

В уравнении заработной платы значимо отличным от нуля на уровне значимости 0.05 оказался коэффициент при дамми переменной на факт проживания в городе. При этом знак данного коэффициента положительный, что соответствует изначальным предположениям о влиянии данной переменной. Так, вероятно, при прочих равных, факт проживания в городе будет повышать заработную плату женщины. Другие коэффициенты и эффект от относящихся к ним переменных не поддаются интерпретации, так как являются

незначимыми, что не позволяет нам на основе данной модели возраст или уровень безработицы в округе влияют на заработную плату женщины.

3.2. Опишите отличия модели Хекмана от модели Тобина.

Во-первых, модель Хекмана, в отличие от модели Тобина, предусматривает, что выбор между альтернативами (вероятность) и наблюдаемый фактор могут описываться разными процессами. Именно эта предпосылка служит причиной того, что в модели Хекмана две латентные переменные, а не одна. Однако из этого следует достаточно жесткая предпосылка о совместном нормальном распределении ошибок. Из первого пункта следует еще одно отличие, которое является преимуществом модели Тобина, — это количество оцениваемых параметров. В модели Тобина их меньше, что повышает эффективность оценок. Еще одним отличием и преимуществом модели Тобина является наличие единственного максимума у функции правдоподобия, что не всегда выполняется в модели Хекмана, оцениваемой через ММП, в случае чего могут наблюдаться проблемы с идентификацией параметров.

3.3. Воспользуйтесь методом Хекмана, основанным на двухшаговой процедуре и сравните оценки, с полученными с использованием метода Хекмана, основанном на методе максимального правдоподобия. Опишите относительные преимущества и недостатки обоих методов.

Результаты оценивания модели Хекмана двухшаговой процедурой представлены в Таблице 4:

Tobit 2 model (sample selection model)				
2-step Heckman / heckit estimation				
753 observations (325 censored and 428 observed)				
12 free parameters (df = 742)				
<u>Probit selection equation:</u>				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.775088	0.725009	7.966	0.00000000000000619 ***
huswage	-0.124772	0.018597	-6.709	0.00000000003885963 ***
mrt	-6.920732	0.906938	-7.631	0.00000000000007171 ***
unem	0.004783	0.015445	0.310	0.757
<u>Outcome equation:</u>				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.95910728	4.70135097	1.055	0.2918
age	0.00857337	0.22291739	0.038	0.9693
I(age^2)	-0.00007988	0.00258941	-0.031	0.9754
city	0.81797970	0.33512569	2.441	0.0149 *
unem	0.02262143	0.05744999	0.394	0.6939

Multiple R-Squared:0.0456, Adjusted R-Squared:0.0342

Error terms:

	Estimate	Std. Error	t value	Pr(> t)
invMillsRatio	-2.6662	0.7811	-3.413	0.000677 ***
sigma	3.7985	NA	NA	NA
rho	-0.7019	NA	NA	NA

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Интерпретация результатов оценивания модели Хекмана двухшаговой процедурой, представленных в *Таблице 4*, позволяет рассуждать аналогично интерпретации *Таблицы 3*, представляющей собой результаты оценивания модели Хекмана ММП. Так как факт значимости коэффициентов и их знаки не отличаются от представленных в *Таблице 3*, подробная интерпретация заново приводиться не будет. Для более наглядного сравнения посмотрим на оценки коэффициентов и корреляции случайных ошибок, полученных обоими способами, представленные в *Таблице 5*:

	Heckman MLE	Heckman 2step
(Intercept)	3.86122202051	4.9591072834
age	0.00735265950	0.0085733683
I(age^2)	-0.00002239166	-0.0000798759
city	0.80026450612	0.8179797032
unem	0.01469033054	0.0226214265
rho	-0.2755582	-0.7019058

Таблица 5. Сравнение двух методов оценивания модели Хекмана.

Как можно видеть по сравнительной Таблице 5, оценки, полученные двухшаговой процедурой, по модулю превосходят оценки ММП. Это может быть связано с отличиями данных способов оценивания.

Преимуществом двухшаговой процедуры и недостатком ММП является возможность получения оценок. Двухшаговый метод работает всегда, в то время как ММП только при наличии у функции правдоподобия единственного максимума.

При этом ММП правдоподобия обладает рядом преимуществ над двухшаговой процедурой. Во-первых, эффективность оценок: ММП оценки более эффективны, что, кстати, может служить причиной того, что они по модулю меньше, то есть, возможно, точнее. Во-вторых, по причине того, что дисперсии у найденных двухшаговым методом оценок могут быть велики, может быть трудно проверять гипотезы. Так, например, незначимость коэффициента может на самом деле быть следствием несовершенства оценки стандартного отклонения. Третьим недостатком двухшагового метода является тот факт, что показывающая смещение из-за неслучайного попадания наблюдений в выборку

функция (обратное отношение Миллса) $\lambda_i = -\frac{\varphi(-z_i^T \gamma)}{1 - \Phi(-z_i^T \gamma)} = \frac{\varphi(z_i^T \gamma)}{\Phi(z_i^T \gamma)}$ на довольно широком участке близка к линейной, что может вызвать мультиколлинеарность. В данном случае необходимы exclusion restrictions, то есть наличие в уравнении отбора уникальных регрессоров, не входящих в уравнение интенсивности.

3.4. Проинтерпретируйте значимость и значение оценки корреляции между случайными ошибками в обоих оцененных моделях. Укажите, можно ли было бы обойтись оценением обычной МНК модели.

Согласно Таблице 3, оценка корреляции между случайными ошибками двух уравнений, полученная при ММП, значимо отлична от нуля при уровне значимости 0.01. При этом ее значение равно -0.2755582. Это может говорить о том, что, вероятно, между случайными ошибками есть значимая отрицательная зависимость. Согласно Таблице 4, иллюстрирующей результаты оценивания двухшаговым методом, о значимости оценки корреляции можно говорить исходя из значимости обратного отношения Миллса. В данном случае оно оказалось значимо отличным от нуля при любом разумном уровне значимости. При этом сама оценка корреляции равна -0.7019058. Это может говорить о том, что, вероятно, между случайными ошибками есть значимая отрицательная зависимость.

Такие результаты могут свидетельствовать о том, что, скорее всего, нельзя было обойтись оценением обычной МНК модели. Данная модель игнорирует смещение отбора, которое возникает по причине неслучайного попадания наблюдений в выборку, из-за чего оценки коэффициентов получаются несостоятельными.

3.5. В любой из двух оцененных в данном разделе моделей для индивида с произвольными характеристиками рассчитайте (предварительно записав формулу):

A) $E(y^|z=1)$ и $E(y^*|z=0)$*

$$E(WAGE_i^* | LFP_i = 1) = x_i^T \beta + \rho \sigma \frac{\varphi(-x_i^T \beta)}{1 - \Phi(-x_i^T \beta)} = x_i^T \beta + \rho \sigma \lambda_i.$$

Характеристики индивида, для которого будут производиться расчеты, представлены в Таблице 2 в Приложении. Помимо упомянутых ранее характеристик, туда также включены доход работающей женщины ($lfp = 1$, $wage = 10$), заработная плата мужа ($huswage = 13$) и налог в округе ($mrt = 0.55$).

Для индивида с характеристиками, представленными в Таблице 2 в Приложении, данная условная вероятность по модели, оцененной ММП, равна 4.528078.

$$E(WAGE_i^* | LFP_i = 0) = x_i^T \beta + \rho \sigma \frac{\varphi(x_i^T \beta)}{1 - \Phi(x_i^T \beta)} = x_i^T \beta + \rho \sigma \eta_i, \text{ где } \eta_i = \frac{\varphi(x_i^T \beta)}{1 - \Phi(x_i^T \beta)}.$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, данная условная вероятность по модели, оцененной ММП, равна 6.040044.

Б) предельный эффект любой переменной (не дамми), входящей линейно и в основное уравнение, и в уравнение занятости, на $E(y^* | z = 1)$ и $E(y^* | z = 0)$.

Переменной, входящей линейно в оба уравнения, является уровень безработицы *unempl*, ее предельный эффект и будет вычислен далее.

$$\begin{aligned} \frac{\partial E(WAGE_i^* | LFP_i = 1)}{\partial unem} &= \beta_{unem} + \rho \sigma \frac{\partial \lambda_i}{\partial unem} = \\ &= \beta_{unem} - \rho \sigma * (x_i^T \gamma \lambda_i + \lambda_i^2) * \gamma_{unem}. \end{aligned}$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, данный предельный эффект (если быть точнее, его оценка) по модели, оцененной ММП, равен 0.01659295.

$$\begin{aligned} \frac{\partial E(WAGE_i^* | LFP_i = 0)}{\partial unem} &= \beta + \rho \sigma \frac{\partial \eta_i}{\partial unempl} = \\ &= \beta_{unem} - \rho \sigma * (x_i^T \gamma \eta_i - \eta_i^2) * \gamma_{unem}. \end{aligned}$$

Для индивида с характеристиками, представленными в *Таблице 2* в *Приложении*, данный предельный эффект (если быть точнее, его оценка) по модели, оцененной ММП, равен 0.01729283.

Часть 4. Модель Ньюи

4.1. Опишите преимущества и недостатки метода Ньюи по сравнению с методом Хекмана.*

Одним из преимуществ метода Ньюи над методом Хекмана является возможность получения состоятельных оценок при маленьком числе допущений за счет аппроксимации полиномом условного математического ожидания. Так, например, в модели Ньюи отсутствует какое-либо допущение о распределении случайных ошибок, в то время как в модели Хекмана выдвигается предпосылка о совместном нормальном распределении случайных ошибок из двух уравнений. Кроме того, метод Ньюи достаточно прост в реализации.

Однако у данного метода есть и недостатки относительно метода Хекмана. Так, по причине аппроксимации только математического ожидания зависимой переменной при условии, что событие произошло, далее мы и можем предсказывать только это математическое ожидание. Еще одним недостатком данной модели является часто возникающая мультиколлинеарность. Это происходит из-за того, что параметры уравнения отбора могут коррелировать с параметрами уравнения интенсивности, а модель их дополнительно дублирует полиномами. Кроме того, в модели Ньюи константа неидентифицируема, то есть очень сложно определить, какая ее часть заложена в свободном члене, а какая – в математическом ожидании случайной ошибки. Это влечет за собой затруднения в предсказании безусловной зависимой переменной. Помимо этого, результаты оценивания не дают явную возможность восстановить совместное распределение случайных ошибок, их корреляцию, в частности, что затрудняет интерпретацию результатов.

Приложение

Таблица 2. Характеристики индивида.

age	35
city	1
unemp	9
work	1
wage	10
huswage	13
mtr	0.55