

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ

Факультет социальных наук
Межфакультетская кафедра общей математики

Эссе по курсу "Анализ категориальных данных"

Различия в эффектах человеческого капитала для индивидов с
различным уровнем дохода на примере исследования испанских
домохозяйств 2020 года

Выполнил: Ластовецкий Дмитрий, БПТ191

Преподаватель: ст.преп. Сальникова Дарья Вячеславовна

Москва 2022

Оглавление

Аннотация	3
Введение	4
Обзор существующей литературы	6
Данные	8
Моделирование при помощи порядковой логит-модели	10
Заключение	13
Список литературы	14

Аннотация

В данной работе мы пытаемся исследовать влияние переменных человеческого капитала на статус бедности, воспроизводя аналогичное и более раннее исследование Филиппа Хонга и Шанты Панди на данных из исследований бюджета испанских домохозяйств (Encuesta de Presupuestos Familiares). Мы оценили влияния образования на вероятность попадания в категории с низким доходом при помощи порядковой логит-модели. Общие выводы модели согласуются с теоретическими положениями: повышение уровня человеческого капитала, измеряемое как уровень образования, действительно снижает вероятность попадания в категории с низким доходом.

Введение

Автор выражает отдельные благодарности:

Хаматдиновой А. - за идею писать об Испании и многолетнюю дружбу

Воскобойникову И.Б. - за поддержку и веру в мои зачаточные академические способности

My work on human capital began with an effort to calculate both private and social rates of return to men, women, blacks, and other groups from investments in different levels of education

Gary Becker

Теория человеческого капитала - одно из основных концептуальных достижений теоретической экономики XX века, которая оказала влияние как на развитие моделирования микроэкономических процессов по типу равновесия на рынке заработных плат [3, Becker, 2009], так и на моделирование долгосрочного экономического роста [8, Romer, 1990]. С точки зрения теории человеческого капитала, инвестиции в человеческий капитал повышают уровень будущей заработной платы [3, Becker, 2009]. Человеческий капитал в формах образования [9, Schultz, 1961][3, Becker, 2009], переобучения [5, Gueron, 2002] и здоровья [6, Grossman, 1972] положительно влияет на производительность труда и экономический рост. В общем, образованные, квалифицированные и здоровые люди обычно зарабатывают больше и имеют больше шансов на восходящую социальную мобильность.

В этой связи логично предположить, что индивиды с низким уровнем человеческого капитала подвергаются большему социальному исключению и не могут претендовать на высокие заработные платы, проживая за чертой бедности. Исследование, результаты которого мы попытаемся воспроизвести [7, Hong & Pandey, 2008] дает количественные оценки таким эффектам: в оцениваемой авторами мультиномиальной логистической регрессии переменные, связанные с человеческим капиталом, действительно положительно влияют на уровень дохода.

Тем не менее, теория человеческого капитала подвергается сегодня серьезной критике: можем ли мы говорить о существенном влиянии человеческого капитала на заработные платы, когда рынок перенасыщен специалистами с высшим образованием, повышение квалификации в среднем общедоступно и социальные

барьеры для людей с низким уровнем здоровья всячески устраняются? В этой связи *актуальна* попытка переоценки модели Хонга и Пэнди на более свежих данных. В оригинальном исследовании авторы используют массив данных исследования дохода и программ участия 1996 года. [1, NBER, 1996]. Мы постараемся переоценить модель на данных 2020 года - из исследования бюджета испанских домохозяйств (Encuesta de Presupuestos Familiares). [2, INE, 2020]. К несчастью, в данной статистике нет данных о дополнительном образовании и здоровье домохозяйств; тем не менее, мы считаем оценку усеченной модели так же востребованной, так как уровень образования в целом является основной переменной человеческого капитала и вносила наибольший вклад в оригинальной модели.

Таким образом, *цель* нашего исследования - проверить, влияет ли уровень человеческого капитала на уровень дохода. *Исследовательский вопрос* состоит в том, является ли значимым влияние уровня образования на уровень дохода в Испании в 2020 году? Используемый нами *метод* - логистическая модель с порядковым откликом. Гипотезы исследования следующие:

- При росте уровня образования вероятность попасть в категорию с более низким доходом уменьшается;
- Уровень образования является более значимым фактором в модели по сравнению с контрольными переменными - гендером, возрастом и семейным положением.

Обзор существующей литературы

В этом разделе мы попытаемся дать теоретическое обоснование того, почему инвестиции в человеческий капитал вообще могут быть связаны с уровнем дохода. Зачем людям инвестировать в переменные человеческого капитала? Достаточно очевидно, что уровень здоровья отражает ожидаемую продолжительность жизни, и, как следствие, является очевидным способом максимизации суммарной жизненной полезности, однако, инвестиции в образование или квалификацию долгое время не были свойственны человечеству в целом.

Так, в рабовладельческой Европе и Америке рабовладельцы не инвестировали в образование своих рабов *вообще* [4, Goldin, 2016], что экономические историки связывают с большими возможностями кооперации и сопротивления рабству у образованных рабов. Так как рабы обычно были представителями разных этносов и только ограниченно знали иностранные языки, они не могли адекватным образом общаться друг с другом и кооперироваться, что могло быть устранено при введении формального образования.

Так, отмена рабовладельческой системы спровоцировала рост обязательного образования в наиболее экономически развитых странах. Начальное образование охватывало более 70% мужского населения от 5 до 14 лет в Германии, Франции и США в середине XIX века [4, Goldin, 2016] и стало *законодательно обязательным* в этих странах. В связи с этим возникает другой экономически значимый вопрос: зачем правительства делают образование обязательным? Почему все семейства не отдали своих детей в учебные заведения без законодательного принуждения, если существуют очевидные выгоды образования?

Наиболее прозрачный ответ предложил Гэри Бекер [3, Becker, 2009]. Согласно его модели, индивиды приобретают образование только в том случае, если при этом их ожидаемый дальнейший доход будет выше, чем в случае отсутствия образования, и покроет альтернативные издержки времени и средств, затраченных на обучение. Так, вместо обучения в университете индивид может работать и получать заработную плату; он согласится обучаться в университете, только если после получения образования его заработная плата будет выше и окупит обучение в университете.

В XIX веке выгоды от получения начального образования не были очевидными, в связи с чем индивиды поступали рационально и просто максимизировали ожида-

емый жизненный доход. Тем не менее, это было не эффективное общественное равновесие: в XIX веке правительства европейских государств уже осознали выгоды промышленной революции, в связи с чем им было необходимо проводить образование населения для работы в промышленности и для возможности изобретения новых технологий в целом [4, Goldin, 2016].

Последующие волны технологических революций и рост симметрии информации между государством и населением только увеличили потребность в образовании, в связи с чем выводы Гэри Бекера в среднем работают и на данных даже конца XX века: у населения действительно есть стимулы получать более высокое образование, так как это увеличивает их потенциальный жизненный доход и покрывает издержки на получение образования [3, Becker, 2009].

Значимым дополнительным вопросом в связи с теорией человеческого капитала является *ловушка бедности* (англ. poverty cycle) [4, Goldin, 2016]. Это стилизованный экономический факт, согласно которому бедность самовозпроизводится из-за отсутствия доступа бедных индивидов к социальным лифтам. В частности, бедным индивидам сложнее получить качественное образование, что не даёт им возможности получить более оплачиваемые должности, что, в свою очередь, не позволяет им дать доступ к качественному образованию у собственных детей. Этот стилизованный факт - основная теоретическая предпосылка нашего исследования: мы предполагаем, что индивиды с низким уровнем дохода не могли получить достаточный уровень образования, что не позволяет им сейчас занимать высокооплачиваемые должности.

Таким образом, оценивание нами модели теоретически обосновано: мы сможем проверить, насколько связан уровень образования и доход.

Данные

Основной источник данных - исследование бюджета испанских домохозяйств (Encuesta de Presupuestos Familiares), проведенное в 2020 году Национальным Статистическим Институтом Испании (Instituto Nacional de Estadística) [2, INE, 2020]. Всего в выборке чуть более 19000 наблюдений. В качестве переменных мы используем:

- *income* - уровень дохода респондента. Порядковая переменная. Изменяется от 1 (менее 500 евро в месяц) до 10 (более 9000 евро в месяц). В рамках исследования мы преобразовали отклик следующим образом

1 - доход респондента менее 1500 евро в месяц, что соответствует двум прожиточным минимумам в Испании (аналогичный показатель использовался в оригинальном исследовании Хонга).

2 - доход респондента от 1500 до 5000 евро, что примерно соответствует средней трети зарплат в Испании.

3 - доход респондента выше 5000 евро.

- *edu* - уровень образования респондента. Порядковая переменная. Изменяется от 1 (респондент не умеет читать или писать или окончил менее 5 лет школьного образования) до 8 (респондент обладает докторской степенью)
- *age* - контрольная целочисленная переменная, изменяется от 1 до 9. Округленное число десятков лет респондента.
- *gender* - контрольная бинарная переменная, обозначающая пол респондента. После преобразования 0 - мужчины, 1 - женщины.
- *ms* - контрольная факторная переменная, обозначающая семейный статус респондента. В преобразованном нами виде 1 - респондент(ка) не состоял(а) в браке; 2 - респондент(ка) состоит в зарегистрированном браке; 3 - респондент(ка) ранее состоял(а) в браке (развод или смерть супруга).
- *child* - контрольная целочисленная переменная, обозначающая число финансово зависимых детей.
- *hisp* - контрольная бинарная переменная, обозначающая наличие испанского

резидентства. В преобразовании 0 - респондент не является резидентом, 1 - респондент является резидентом.

Посмотрим на описательные характеристики наблюдений:

Таблица 1: Описательные статистики выборки

income	edu	age	gender	ms	child	hisp
1:3007	2 :2721	Min. :0	0:12784	1: 3783	Min. :0	0: 57
2:9231	3 :5488	1st Qu.:2	1: 6386	2:11243	1st Qu.:0	1:19113
3:6932	4 :3577	Median :5		3: 4144	Median :0	
	5 :2076	Mean :4.49			Mean :0.6	
	6 :1768	3rd Qu.:7			3rd Qu.:1	
	7 :2933	Max. :9			Max. :9	
	(Other): 607					

Выборка достаточно сбалансирована по всем переменным, кроме *hisp*. Это можно связать с характером данных: данные являются опросными, а вопрос о наличии статуса резидента является достаточно сензитивным. Тем не менее, это достаточно репрезентативные данные: действительно, в Испании мало нелегальных иммигрантов.

К несчастью, плохо сбалансированы и наблюдения по отклику *income*. Однако, это теоретически обоснованно: в Испании достаточно развиты системы государственной поддержки, что уменьшает число потенциальных респондентов с низким уровнем дохода.

Моделирование при помощи порядковой логит-модели

Для моделирования мы использовали порядковую логит-модель. В качестве отклика выступала категория дохода респондента, в качестве контрольных переменных - возраст, гендер, семейное положение, количество детей и наличие статуса резидента Испании, в качестве объясняющей переменной - категория, соответствующая уровню образования. Модель показывает отличные результаты при тесте Хосмера-Лемешова ($\chi^2 = 151.91$; $p_{value} < 0.001$), что в целом означает, что модель с предикторами предсказывает отклик лучше, чем пустая модель. В нижеследующих таблицах приведены оценки параметров модели:

Таблица 2: Коэффициенты при предикторах

	Value	Std. Error	t value
edu	0.4996	0.009436	52.95
age	0.004102	0.005111	0.8026
gender	-0.7913	0.03183	-24.86
ms	-0.006714	0.0232	-0.2894
child	0.2078	0.01648	12.61
hisp	-0.2979	0.2656	-1.122

Таблица 3: Intercepts

	Value	Std. Error	t value
1 2	-0.3277	0.2732	-1.2
2 3	2.327	0.2736	8.504

Для удобства интерпретации преобразуем оценки в терминах отношения шансов:

Таблица 4: отношения шансов

edu	age	gender	ms	child	hisp
1.648	1.004	0.4533	0.9933	1.231	0.7424

Модель показывает, что, действительно, образование оказывает существенный эффект на доход. Так, при росте образовательной степени на единицу, шансы попасть в категорию с более высоким доходом в среднем увеличиваются примерно на 65%. В то же время, остальные предикторы, кроме гендера, оказывают меньшее влияние: респонденты-мужчины в среднем на примерно 65% реже попадают в категорию с более низким доходом. Проверка коэффициентов на значимость тестом Вальда показывает, что коэффициент при переменной *edu* статистически отличен от нуля ($\chi^2 = 1208.3$; $p_{value} < 0.001$).

Более детальный анализ модели с составлением confusion matrix показывает, что модель часто неверно предсказывает респондентов с наиболее низким уровнем дохода:

Таблица 5: Confusion matrix

	Reference 1	Reference 2	Reference 3
Predicted 1	122	56	6
Predicted 2	2584	7139	3435
Predicted 3	301	2036	3491

Таблица 6: Метрики качества confusion matrix

	Sensitivity	Specificity	Balanced Accuracy
Class 1	0.04057	0.9962	0.5184
Class 2	0.7734	0.3944	0.5839
Class 3	0.5036	0.809	0.6563

О том, как можно попробовать решить эту проблему, мы попытаемся сказать

в заключении. На данный момент кратко скажем, что, возможно, группировка респондентов по уровню дохода, предложенная в оригинальной статье, была не самой удачной, и необходима более точная группировка и более точные данные. Также можно сказать, что возможно, необходимы дополнительные преобразования предиктора или попытки искусственно сбалансировать выборку.

Тем не менее, модель обладает неплохими характеристиками общей точности ($Accuracy = 0.5609$ с доверительным интервалом $[0.5538; 0.5679]$ против 0.4815 у пустой модели).

Таким образом, уровень образования действительно оказывает положительный эффект на доход.

Заключение

В предложенной работе мы показали, что рост уровня образования действительно снижает вероятность оказаться в категории с более низким доходом, что подтверждает общие теоретические предположения теории человеческого капитала и согласуется с предположением о положительной отдачей от образования. Модель показала, что образование является одним из наиболее важных факторов при объяснении разброса уровня дохода, более важным, чем стандартные контрольные факторы для моделей, объясняющих неравенство.

Тем не менее, наша модель столкнулась с рядом проблем: она неудовлетворительно прогнозирует результаты для индивидов с наиболее низким доходом. Для устранения этой проблемы мы можем предложить следующие варианты:

- Перегруппировать население иначе в соответствии с уровнем дохода. Можно оценить стандартную линейную модель на изначальных категориях респондентов или предложить другую группировку, возможно, с большим или меньшим числом категорий.
- Предложить другой характер связи между откликом и предиктором. К примеру, использовать в качестве предиктора квадрат числа полных лет образования.
- Последовать логике оригинального исследования и переоценить модель с бинарным откликом (где отклик - вероятность оказаться в категории с доходом ниже уровня бедности) и мультиномиальным откликом (в предположении об отсутствии отношений порядка между категориями отклика)

Исследование также можно расширить, введя панельную структуру данных и анализируя модели с включением временной динамики, или введя дополнительные иерархические единицы, к примеру, регионы Испании как переменные более высокого уровня.

Список литературы

- [1] Survey of Income and Program Participation (SIPP). Date Views: 01.06.2022. URL: www.nber.org/research/data/survey-income-and-program-participation-sipp
- [2] Instituto Nacional de Estadística. Date Views: 01.06.2022. URL: <https://ine.es/>.
- [3] Becker, G. S. (2009). Human capital: A theoretical and empirical analysis, with special reference to education. University of Chicago press.
- [4] Goldin, Claudia. 2016. "Human Capital." In Handbook of Cliometrics, ed. Claude Diebolt and Michael Hauptert, 55-86. Heidelberg, Germany: Springer Verlag.
- [5] Gueron, J. M., & Hamilton, G. (2002). The role of education and training in welfare reform. Washington, DC: Brookings Institution.
- [6] Grossman, M. (1972) On the concept of health capital and the demand for health. J. Polit. Econ. 80(2), 223–255.
- [7] Hong, P. Y. P., & Pandey, S. (2008). Differential effects of human capital on the poor and near-poor: Evidence of social exclusion. Journal of Poverty, 12(4), 456-480.
- [8] Romer, P. M. (1990). Endogenous technological change. Journal of political Economy, 98(5, Part 2), S71-S102.
- [9] Schultz, T.W. (1961). Investment in human capital. American Economic Review, 161:1-17.