

ОТЧЕТ

В данной работе были построены несколько моделей, предсказывающих изменение частоты посещения респондентом ресторанов и баров в случае запрета курения в них. По итогу работы были сделаны выводы о том, какая из моделей является более предпочтительной.

В качестве зависимой переменной в каждой модели выступала переменная visit, которая характеризует направление изменения частоты посещения ресторанов и баров после введения запрета на курение в них: 1 – если респондент станет реже ходить в рестораны и бары, 2 – если частота посещения никак не изменится, 3 – если респондент станет чаще ходить в данные заведения. В таблице 1 представлены размеры соответствующих категорий.

Номер категории	Количество наблюдений, шт	Доля, %	Нарастающим итогом, %
№ 1	81	6.69	6.69
№ 2	990	81.75	88.44
№ 3	140	11.56	100
Итого	1211	100	100

Таблица 1. Данные о численности категорий

Как видно, наибольшее количество наблюдений приходится на 2 категорию (81.75%). Крайние категории являются относительно малочисленными, что может привести к проблемам при их прогнозировании (особенно тех, кто после запрета станет реже ходить). При этом, медианным представителем в данном случае является респондент, поведение которого в отношении посещения ресторанов и баров никак не изменится (он же является типичным представителем выборки с позиции зависимой переменной).

В таблице 2 представлены факторы, которые могут влиять на частоту посещения респондентами ресторанов и баров в случае введения вышеупомянутого запрета.

Название переменной	Принимаемые значения	Фактор	Описание
male	1 – мужчина 0 – женщина	Пол	Можно предположить, что у мужчин и женщин наблюдается разная интенсивность и переносимость курения. Это означает, что запрет курения в барах и ресторанах изменит частоту посещения мужчинами данных заведений в большей или меньшей степени, чем частоту посещения женщинами. Однако, само направление предугадать сложно (
age	Положительные целые числа	Возраст	С увеличением возраста люди все в меньшей степени располагают свободным временем, а значит тщательнее выбирают места, чтобы провести там свое время. Изменение условий посещения ресторанов и баров у взрослых людей в большей степени изменит предпочтения в их посещении, чем у относительно молодых.

income	Положительные числа	Доход на человека в домохозяйстве респондента	Люди с высоким доходом могут позволить себе больший выбор мест, которые они могут посетить. Следовательно, при изменении условий посещения баров и ресторанов (запрет курения), люди с более высоким доходом в большей степени подвержены изменениям в своих предпочтениях относительно их посещений. Однако, само направление предугадать сложно.
smoke	1 – курит 0 – не курит	Курение	Запрет курения в барах и ресторанах может негативно сказаться на частоте посещения этих мест людьми, которые курят и наоборот, увеличит частоту посещения у некурящих людей.
health	1 – да, слежу за здоровьем 0 – нет, не слежу за здоровьем	Внимание к здоровью	Запрет курения в барах и ресторанах может повысить посещаемость этих заведений людьми, которые следят за своим здоровьем.

Таблица 2. Описание переменных.

Как видно из таблицы, для приведенных факторов, определить однозначное направление влияния достаточно сложно, поскольку оно может зависеть от структуры выборки (например, от соотношения тех, кто курит и тех, кто не подвержен данной привычке). В таблице 3 приведена информация о различиях по этим факторам групп, каждая из которых характеризует поведение респондентов в случае ввода запрета на курение в ресторанах и барах.

Фактор	Номер категории зависимой переменной					
	1		2		3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Пол	0.44	0.50	0.39	0.49	0.27	0.45
Возраст	35.35	14.72	36.75	14.44	35.72	13.92
Доход на человека в домохозяйстве респондента	11249.32	8786.21	13003.89	10439.30	13265.06	9027.47
Курение	0.62	0.49	0.37	0.48	0.09	0.28
Внимание к здоровью	0.57	0.50	0.61	0.49	0.84	0.37

Таблица 3. Сопоставление групп зависимой переменной по факторам

Как видно из таблицы, по некоторым факторам данные группы существенно отличаются. Так, если в первой категории доля курящих людей составляет 62%, то во второй только 37%, а в третьей всего лишь 9%. Тенденция на уменьшение доли по мере перехода из первой категории в третью наблюдается и по переменной пола. Так, если в первой категории, доля мужчин составляет 44%, то в третьей уже 27%. Аналогичная тенденция наблюдается и по вниманию к здоровью: при переходе из 1 категории в 3, доля людей, следящих за своим здоровьем, увеличивается с 57% до 84%. Наконец стоит отметить, что среднее значение дохода в 3 категории выше, чем в 1 и почти такое же, как во 2 категории. Что касается возраста, то его среднее значение от категории к категории слабо различается (в пределах 1,5 лет).

Для объяснения зависимой переменной, отвечающей за частоту посещения ресторанов и баров после введения запрета на курение в них с использованием рассмотренных выше факторов была оценена логит модель упорядоченного выбора, оценки которой представлены в таблице 4 (здесь и далее обозначение значимости коэффициентов следующее: * $p < .1$; ** $p < .05$; *** $p < .01$)

Зависимая переменная: частота посещения ресторанов и баров (1 – минимум, 3 – максимум)	Оценки
Пол	-0.0840 (0.162)
Возраст	-0.00457 (0.00526)
Доход на человека в домохозяйстве респондента	1.32e-05* (7.07e-06)
Курение	-1.393*** (0.196)
Внимание к здоровью	0.517*** (0.165)
Граница между 1 и 2 категориями	-3.105*** (0.300)
Граница между 2 и 3 категориями	2.036*** (0.277)
Расчетная статистика	LR chi2(5)=89.94
p-value	0.0000

Таблица 4. Оценка логит модели

Как видно из таблицы, значимыми оказались следующие факторы: доход, курение и внимание к здоровью. Глядя на оцененные коэффициенты, мы не можем сказать, как будет вести себя вероятность того, что человек не изменит своего предпочтения в отношении посещения ресторанов и баров. Однако, мы можем сказать, глядя на знак коэффициента, как будет меняться вероятность по мере перехода из категории 1 в категорию 3. Так, увеличение дохода сопровождается уменьшением вероятности того, что респондент начнет ходить реже в рестораны и бары после введения запрета (хотя значение коэффициента близко к нулю, на 10% уровне он значим). Внимание к здоровью также сопровождается уменьшением вероятности того, что респондент начнет ходить реже в рестораны и бары. Наконец, факт курения респондентом как минимум негативно сказывается на вероятности, что респондент будет ходить чаще в данные заведения после введения запрета на курение. Пол и возраст в оцененной модели оказались незначимыми. Тест отношения правдоподобия показывает, что модель в целом значима (высокий h^2 , как следствие, нулевой p-value).

Стоит отметить, что помимо оценок коэффициентов, упорядоченная логит модель оценивает и границы для латентной переменной, с помощью которых то или иное наблюдение может быть отнесено к соответствующей категории. В данном случае, доверительные интервалы оценок границ не пересекаются (оценки границ достаточно далеко расположены друг от друга, а значения среднеквадратических отклонений не являются большими), что является положительным сигналом, поскольку модель четко может определить границы соответствующих категорий и верно осуществить классификацию.

В таблице 5 представлена информация о средних предсказанных вероятностях каждого исхода и долях респондентов, выбравших эти исходы на самом деле.

Номер категории	Количество наблюдений, штук	Средние предсказанные вероятности, %	Фактическая доля, %
№ 1	1211	6.64	6.69
№ 2	1211	81.71	81.75
№ 3	1211	11.65	11.56

Таблица 5. Сопоставление предсказанных вероятностей и фактических долей

Из таблицы можно сделать вывод, что средние предсказанные вероятности оцененной модели практически совпадают с долями респондентов, которые фактически выбрали ту или иную категорию (но это не означает, что оцененная модель верно классифицирует конкретного респондента в фактически выбранную им категорию). Данная таблица говорит, что предсказания вероятностей отнесения респондентов к той или иной категории верно соотносится со структурой выборки, по которой оценивалась модель.

Одним из наиболее значимых факторов в оцененной модели оказалось курение. При этом очевидно, что предпочтения курящих и не курящих людей относительно запрета курения в барах и ресторанах может существенно отличаться между собой. Это связано с тем, что курящий человек создает отрицательный внешний эффект для некурящего. Поэтому, запрет курения уменьшает полезность курящих людей в ресторанах и барах и увеличивает ее для некурящих за счет устранения отрицательной экстерналии. Такая разнонаправленность эффектов от запрета курения, которая может проявляться и через другие регрессоры, требует проверить гипотезу о необходимости оценки отдельных моделей для курящих и некурящих респондентов с помощью теста отношения правдоподобия, результаты которого представлены в таблице 6.

Показатель	Значение
LR chi2(6)	72.82
p-value	0.0000

Таблица 6. Результаты теста на отдельные модели

Как видно, на любом допустимом уровне значимости результаты теста говорят, что как для курящих, так и для некурящих, для прогнозирования изменения частоты посещения респондентом ресторанов и баров при вводе запрета на курение требуются отдельные модели. Поэтому, дальнейший анализ будет проводиться уже для двух отдельных моделей, а не для одной.

Еще одним значимым регрессором в первоначально оцененной модели был доход на человека в домохозяйстве респондента. Вместе с тем, вполне возможно, что влияние дохода на латентную переменную может быть нелинейным (существование точки насыщения, при достижении которой, направление влияния дохода может измениться). Следовательно, для проверки гипотезы о нелинейности влияния дохода на латентную переменную, данный фактор был включен в отдельные модели для курящих и не курящих респондентов с квадратом. В таблице 7 приведены оценки отдельных моделей для курящих и не курящих респондентов, при этом в одном случае предполагалось, что доход оказывает линейное влияние на латентную переменную, а в другом, что влияние является нелинейным.

Зависимая переменная: частота посещения ресторанов и баров (1 – минимум, 3 – максимум)	Модели для курящих		Модели для некурящих	
	с линейным доходом	с нелинейным доходом	с линейным доходом	с нелинейным доходом
Пол	0.104 (0.281)	0.109 (0.281)	-0.245 (0.204)	-0.281 (0.205)
Возраст	0.0173 (0.0119)	0.0175 (0.0119)	-0.0101* (0.00597)	-0.0104* (0.00601)

Доход на человека в домохозяйстве респондента	7.03e-06 (1.30e-05)	-4.28e-06 (2.89e-05)	1.58e-05* (8.52e-06)	5.86e-05*** (1.91e-05)
Внимание к здоровью	0.427 (0.280)	0.429 (0.280)	0.565*** (0.207)	0.526** (0.208)
Квадрат дохода на человека в домохозяйстве респондента		2.02e-10 (4.61e-10)		-7.45e-10** (2.95e-10)
Граница между 1 и 2 категориями	-1.097** (0.483)	-1.179** (0.519)	-3.108*** (0.351)	-2.836*** (0.367)
Граница между 2 и 3 категориями	4.556*** (0.600)	4.477*** (0.625)	1.818*** (0.319)	2.129*** (0.346)
Расчетная статистика	LR chi2(4)=5.01	LR chi2(4)=5.20	LR chi2(4)=16.00	LR chi2(4)=22.08
p-value	0.2865	0.2675	0.0030	0.0002
Количество наблюдений	428	428	783	783

Таблица 7. Оценки моделей отдельных для курящих и некурящих респондентов

Как видно, ни одна из приведенных моделей для курящих не является значимой (низкий h^2 , как следствие, высокий p-value). В то же время, обе модели для некурящих респондентов являются значимыми. Кроме того, в модели для некурящих людей подтвердилась гипотеза о нелинейном влиянии дохода на латентную переменную (коэффициент при квадрате дохода является значимым на 5% уровне).

Что касается незначимости моделей для некурящих людей, то скорее всего можно утверждать, что подобранные регрессоры либо действительно слабо объясняют поведение курящих при введении запрета на курение, либо действительно, такой запрет слабо влияет на поведение таких людей. Если второй вариант верен, то можно считать, что полезность от посещения ресторанов и баров для курильщиков, где нельзя курить, выше, чем отказ от посещения этих мест. Кроме того, в реальной жизни практикуется и такое, что курящие люди не отказываются от посещения мест, где запрещено курить, поскольку всегда есть возможность выйти на улицу или в специально отведенные места для курения. Таким образом, если бы был регрессор, который характеризовал бы наличие или отсутствие таких специально отведенных мест возле ресторанов и баров (или возможности выйти на улицу, чтобы безнаказанно покурить), то модель для курящих вполне могла бы быть значимой. Ввиду отсутствия таких регрессоров, а также из содержательных соображений, далее в анализе будет рассматриваться только модель для некурящих людей.

Как видно из таблицы 7, по сравнению с первоначальной моделью стал значим фактор, отвечающий за возраст. При этом, для некурящих людей при увеличении возраста снижается вероятность того, что они будут чаще ходить в рестораны и бары при введении запрета на курение. По-прежнему незначимым регрессором является пол. Другими словами, в данном случае первоначальное предположение о разной переносимости курения между мужчинами и женщинами может быть отвергнуто. Для определения возможности удаления данного регрессора из модели, была построена ограниченная модель, которая не включала в качестве регрессора пол. Далее, для тестирования ограничений, был использован LR-тест, результаты которого приведены в таблице 8.

Показатель	Значение
LR chi2(1)	1.90
p-value	0.1678

Таблица 8. Результаты теста на незначимость пола

Как видно, нулевая гипотеза не отвергается. То есть модель с ограничениями не вложена в модель без ограничений. Следовательно, необходимо предпочесть ограниченную модель, в которой отсутствует незначимый регрессор.

Еще одной причиной, по которой ограниченную модель (без учета пола) стоит предпочесть полной (с учетом пола) являются значения их информационных критериев, которые приведены в таблице 9 вместе с оценками данных моделей.

Зависимая переменная: частота посещения ресторанов и баров (1 – минимум, 3 – максимум)	Полная модель (с учетом пола)	Ограниченная модель (без учета пола)
Пол	-0.281 (0.205)	
Возраст	-0.0104* (0.00601)	-0.0099* (0.006)
Доход на человека в домохозяйстве респондента	5.86e-05*** (1.91e-05)	0.0000554*** (0.000019)
Внимание к здоровью	0.526** (0.208)	0.513** (0.207)
Квадрат дохода на человека в домохозяйстве респондента	-7.45e-10** (2.95e-10)	-7.17e-10** (2.94e-10)
Граница между 1 и 2 категориями	-2.836*** (0.367)	-2.773*** (0.363)
Граница между 2 и 3 категориями	2.129*** (0.346)	2.182*** (0.344)
Расчетная статистика	LR chi2(4)=22.08	LR chi2(3)=20.18
p-value	0.0002	0.0002
AIC	937.0395	936.9422
BIC	965.0183	960.2579

Таблица 9. Оценки полной и ограниченной модели

Из таблицы следует, что значение информационных критериев ограниченной модели хоть и незначительно, но меньше, чем у полной. Это означает, что последняя немного уступает в качестве модели, которая не учитывает пол.

Как уже говорилось ранее, LR тест показал, что для курящих и некурящих респондентов необходимы отдельные модели. Поскольку модель для курящих оказалась незначимой, выбор был остановлен только на модели для некурящих людей. Так как проверка гипотезы о нелинейном влиянии дохода на латентную переменную в данной модели была подтверждена, то нелинейность дохода была учтена. Вместе с тем, из модели был исключен такой фактор, как пол: тест отношения правдоподобия, а также значения информационных критериев показали, что учет пола в модели является лишним. В силу вышеизложенных фактов дальнейший анализ проводился для ограниченной модели, оценки которой были представлены в таблице 9.

Поскольку для интерпретации упорядоченной логит модели используются предельные эффекты, которые для каждого конкретного респондента принимают свои значения, необходимо выбрать такого представителя выборки, предельные эффекты которого наилучшим образом могут служить для интерпретации модели в целом. Для

описания типичного представителя выборки можно взять такие показатели как мода или медиана. Мода описывает такого представителя, который чаще всего встречается в выборке. Однако, в случае модели упорядоченного выбора, такой показатель является не самым лучшим вариантом, поскольку наиболее часто встречающийся респондент может попасть в крайнюю категорию и тогда теряется значительная информация о выборке и переходе между категориями. Медиана описывает такого представителя, который делит всю выборку на две равные части и, таким образом, более полно характеризует выборку в целом. Поскольку в данном случае рассматриваемая модель оценена только по некурящим респондентам, то использование таблиц 1 и 3 для определения типичного (медианного) представителя не представляется возможным. В таблице 10 представлена информация о выборке, состоящей из некурящих представителей первоначальной выборки (пол в данном случае не рассматривается, поскольку в модели он не учитывается).

Фактор	Среднее значение регрессора в каждой из категории зависимой переменной		
	№1	№2	№3
Возраст	40.32	38.05609	35.55469
Доход на человека в домохозяйстве респондента	9504.13	12231.68	13547.07
Внимание к здоровью	0.87	0.67	0.87
Количество наблюдений, шт	31	624	128
Доля, %	3.96	79.69	16.35
Нарастающим итогом, %	3.96	83.65	100

Таблица 10. Данные о численности категорий новой выборки и средних значениях регрессоров

Согласно таблице 10, типичным (медианным) представителем с точки зрения зависимой переменной является респондент, который попал во 2 категорию (именно на эту категорию выпадает граница, которая делит всю выборку на 2 равные части). При этом, средний возраст типичного респондента составляет 38 лет, он следит за своим здоровьем (среднее значение по данному регрессору выше 0.5), а средний доход на человека в его домохозяйстве составляет 12 231 руб.

На рисунке 1 представлены предельные эффекты по доходу для типичного представителя в интервале от минимального (400 руб.) до максимального (100 000 руб.) дохода с шагом 2 000 руб. для вероятности того, что респондент будет чаще посещать рестораны и бары. В приложении 1 приведены значения рассчитанных эффектов.

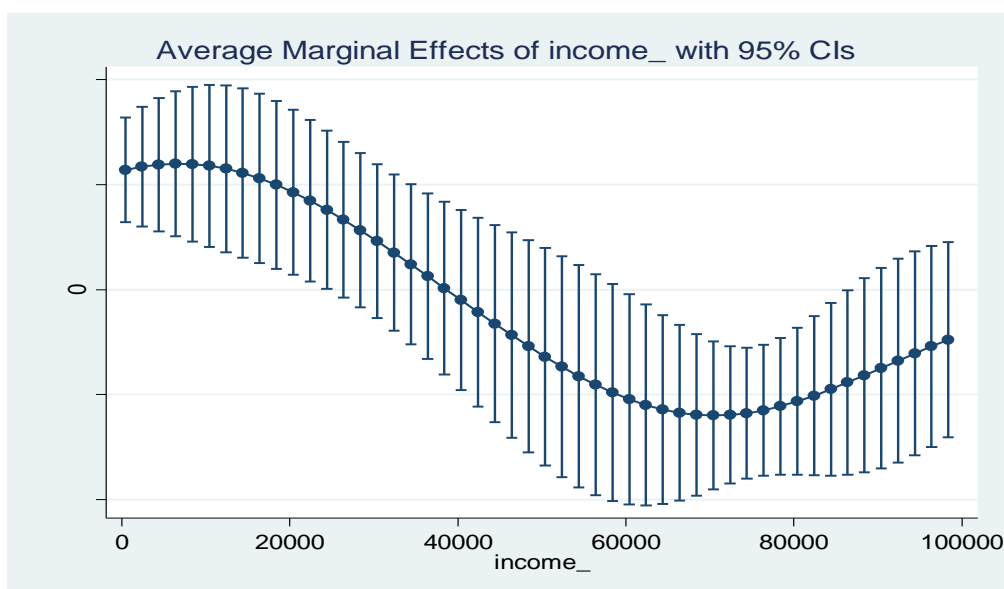


Рисунок 1. График предельных эффектов по доходу для типичного представителя выборки

Как видно из рисунка и приложения 1, предельные эффекты по доходу ведут себя по-разному в зависимости от размера дохода. К примеру, пока средний доход на одного человека в домохозяйстве типичного респондента меньше 8 400 руб., предельный эффект положителен и возрастает. При достижении такого значения дохода, он начинает падать (однако все же остается положительным) и при доходе в 28 400 руб. перестает быть значимым (другими словами, его значение близко к нулю). Предельный эффект снова приобретает значимость при доходе в 56 400 руб., однако в этой точке он уже имеет отрицательный знак. Наконец, при доходе в 90 400 руб. он снова становится незначимым.

На рисунке 2 представлен график динамики вероятности того, что типичный представитель будет чаще посещать рестораны и бары в зависимости от дохода.

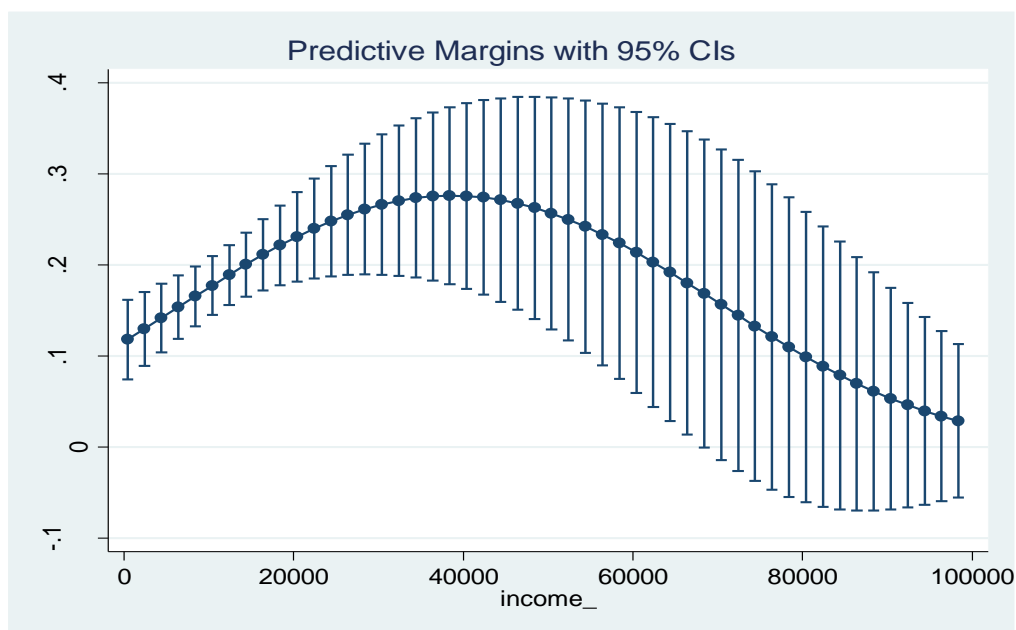


Рисунок 2. График динамики вероятности для типичного представителя выборки

Отличием данного графика от предыдущего является то, что в данном случае по оси ординат отложены сами вероятности увеличения частоты посещения, в то время как на рисунке 1 на оси ординат отложены предельные эффекты по доходу (насколько процентов изменится вероятность при конкретном доходе, если величина дохода увеличится).

Как видно из рисунка, доход оказывает нелинейное влияние на вероятность увеличения посещений ресторанов и баров типичным представителем. Это обусловлено сменой знака предельного эффекта по доходу. Стоит отметить, что наибольшая вероятность того, что типичный представитель будет чаще посещать рассматриваемые заведения (27.61%) наблюдается при уровне дохода в 38 400 руб. (область, где значение предельного эффекта близко к нулю). До достижения этого уровня дохода вероятность увеличения частоты посещения растет, однако по его достижению вероятность начинает падать. Вполне возможно, что поскольку изначально, по мере роста дохода становятся доступны более частые посещения ресторанов и баров, то с запретом курения вероятность увеличения частоты посещения некурящими людьми по мере роста дохода будет увеличиваться. При превышении указанного выше дохода вполне возможно, что некурящие люди перестают быть чувствительны к запрету курения и не увеличивают частоту своего посещения, поскольку просто располагают меньшим свободным временем (высокий доход должен быть чем-то обусловлен), либо имеют другие увлечения и интересы (следовательно, и до ввода запрета редко посещали рестораны и бары), которые не могут позволить себе люди с меньшими доходами.

Оцененная модель предполагает постоянство коэффициентов при переходах между категориями. Вместе с тем, данная предпосылка о параллельности линий может не выполняться и тогда необходимо оценить обобщенную логит модель. В таблице 11 представлены результаты тестирования гипотезы о том, что коэффициенты являются постоянными.

Тест	Chi2	df	p-value
Woldfe Gould	21.22	3	0.000
Brant	15.82	3	0.001
Score	18.66	3	0.000
Likelihood ration	21.22	3	0.000
Wald	16.02	3	0.003

Таблица 11. Результаты проверки гипотезы о необходимости оценки обобщенной логит модели

Как видно из таблицы, все проведенные тесты отвергают нулевую гипотезу о постоянстве всех коэффициентов (иными словами, хотя бы один коэффициент при переходах между категориями не является постоянным). Следовательно, возникает необходимость оценивания обобщенной логит модели, которая учитывает то, что предположение о параллельности линий не работает. В таблице 12 представлены ее оценки.

Зависимая переменная: частота посещения ресторанов и баров (1 – минимум, 3 – максимум)	Переходы между категориями	
	Между 1 и 2	Между 2 и 3
Возраст	-0.00966 (0.0110)	-0.00926 (0.00669)
Доход на человека в домохозяйстве респондента	0.000123*** (4.61e-05)	4.75e-05* (2.65e-05)
Внимание к здоровью	-1.167** (0.544)	1.057*** (0.275)
Квадрат дохода на человека в домохозяйстве респондента	-1.31e-09** (5.18e-10)	-6.91e-10 (5.37e-10)
Граница	3.442*** (0.765)	-2.553*** (0.421)
Расчетная статистика	LR chi2(6)=41.41	
p-value	0.0000	

Таблица 12. Результаты оценки обобщенной логит модели

В данном случае обобщенная логит модель является значимой (высокий h^2 , следовательно, низкий p-value). При этом, данная модель оценивает теперь два блока коэффициентов: то есть теперь при переходе между 1 и 2 категорией и переходе между 2 и 3 категориями оцененные значения коэффициентов отличаются между собой (как было показано ранее, гипотеза о постоянстве всех коэффициентов была отвергнута). Стоит отметить, что существуют отличия и в значимости коэффициентов, оцененных для того или иного перехода. Кроме того, по сравнению с обычной логит моделью, полностью потерял значимость возраст.

Для перехода между 1 и 2 категориями значимыми оказались внимание к здоровью, а также нелинейное влияние дохода на человека в домохозяйстве респондента. Так, если некурящий человек начинает уделять внимание своему здоровью, то вероятность того, что

он начнет реже ходить в рестораны и бары после запрета в них курения, будет уменьшаться по сравнению с вероятностью того, что этот человек не изменит своего поведения в отношении частоты посещения данных мест, либо того, что он станет посещать их чаще. Направление дохода в данном случае проинтерпретировать достаточно сложно, поскольку имеется нелинейный эффект. Можно сказать, что когда доход является относительно мал, то при его увеличении сначала будет наблюдаться и увеличение вероятности неизменности или роста частоты посещения данных мест по сравнению с вероятностью того, что человек станет реже посещать рестораны и бары. Однако, по достижении какого-то уровня дохода, эти различия в вероятностях начнут сокращаться.

Для перехода между 2 и 3 категориями значимыми оказались внимание к здоровью, а также линейное влияние дохода на человека в домохозяйстве респондента. Так, если человек начнет более пристально следить за своим здоровьем, то вероятность того, что он начнет чаще посещать рестораны и бары вырастет по сравнению с вероятностью того, что он либо не изменит частоту своего посещения данных заведений, либо станет реже в них ходить. В случае дохода наоборот, при его увеличении вероятность того, что человек начнет чаще ходить в данные заведения, упадет по сравнению с вероятностью того, что некурящий респондент уменьшит или не изменит частоту своего посещения ресторанов и баров.

Как было указано ранее, гипотеза о постоянстве всех коэффициентов была отвергнута. Однако, это не означает, что каждый коэффициент является непостоянным. Поэтому, для выявления в модели факторов, которые действительно отличаются от категории к категории была использована гамма-параметризация. Отличия между коэффициентами представлены в таблице 13.

Зависимая переменная: частота посещения ресторанов и баров (1 – минимум, 3 – максимум)	Различия между коэффициентами	Переходы между категориями	
	Оценка	Между 1 и 2	Между 2 и 3
Возраст	0.0004032 (0.0123132)	-0.00966 (0.0110)	-0.00966 (0.0110)
Доход на человека в домохозяйстве респондента	-0.0000755 (0.0000515)	0.000123*** (4.61e-05)	0.000123*** (4.61e-05)
Внимание к здоровью	2.223603*** (0.5987327)	-1.167** (0.544)	1.057*** (0.275)
Квадрат дохода на человека в домохозяйстве респондента	6.19e-10 (7.21e-10)	-1.31e-09** (5.18e-10)	-1.31e-09** (5.18e-10)
Граница	-	3.442*** (0.765)	-2.553*** (0.421)
Расчетная статистика	-	LR chi2(6)=41.41	
p-value	-	0.0000	

Таблица 13. Значимость отличий между коэффициентами

Из таблицы видно, что единственным коэффициентом, по которому различия являются значимыми выступает коэффициент при факторе, отвечающем за внимание к здоровью. Это означает, что только по этому фактору значения коэффициентов при переходе между категориями будут отличаться между собой. По всем остальным факторам наблюдается постоянство коэффициентов.

Поскольку гипотеза об отсутствии постоянства всех коэффициентов подтвердилась, стоит предпочесть обобщенную логит модель. При этом, использование гамма параметризации показало, что значимые различия наблюдаются только по коэффициенту при факторе, отвечающем за внимание к здоровью. Таким образом, наилучшей моделью в данном случае является обобщенная логит модель, оценки которой представлены в таблице 13.

Таким образом, были оценены несколько упорядоченных моделей, предсказывающих вероятность того, что респондент будет реже, также или чаще посещать рестораны и бары после введения запрета на курение. Выбор был остановлен на обобщенной логит модели упорядоченного выбора с использованием гамма-параметризации. Данная модель была оценена по выборке, состоящей только из некурящих людей, поскольку результаты теста отношения правдоподобия показали, что для курящих и некурящих людей требуются отдельные модели (можно это интерпретировать и как косвенный признак того, что факт курения действительно влияет на поведение респондентов в отношении частоты посещения, поэтому первоначальное подтверждение о значимости данного фактора подтвердилось). Отдельная модель для курящих людей с тем составом регрессоров, который первоначально был определен, призналась в целом не значимой. Следовательно, все первоначальные предположения о курящих людях не подтвердились. Вполне возможно, что включив в модель регрессор, который бы отвечал за наличие специально отведенных мест для курения у баров и ресторанов, а также потенциальную возможность безнаказанно покурить на улице, выйдя из данных заведений, сделал бы модель значимой (и сделал бы возможным продолжить анализ относительно курящих людей). Отдельная модель для некурящих людей оказалась значимой. При этом, из итоговой спецификации оцененной модели можно сделать выводы о подтверждении или опровержении первоначальных результатов относительно некурящих людей. Не подтвердились предположения о различном влиянии запрета курения в барах и ресторанах для мужчин и женщин (для некурящих), поскольку соответствующий коэффициент в модели оказался незначимым. Различия в возрасте также никак не влияют на поведение. Подтвердилось предположение о значимости дохода, однако сам механизм влияния оказался достаточно сложным (нелинейная зависимость). Наконец относительно такого фактора, как внимание к здоровью, предположения полностью подтвердились: фактор оказался значимым. Кроме того, если респондент начнет более пристально следить за своим здоровьем, то действительно, вероятность того, что он начнет чаще ходить в рестораны и бары, возрастет.

Приложение 1. Расчеты предельных эффектов по доходу в интервале от минимального до максимального дохода с шагом 2 000 руб. для типичного представителя выборки

Значение дохода	dy/dx	Std. Err.	p-value
400.00	5.70e-06	1.27e-06	0.000
2400.00	5.86e-06	1.45e-06	0.000
4400.00	5.95e-06	1.62e-06	0.000
6400.00	5.99e-06	1.77e-06	0.001
8400.00	5.97e-06	1.88e-06	0.002
10400.00	5.89e-06	1.97e-06	0.003
12400.00	5.75e-06	2.03e-06	0.005
14400.00	5.55e-06	2.05e-06	0.007
16400.00	5.30e-06	2.06e-06	0.010
18400.00	4.99e-06	2.04e-06	0.014
20400.00	4.63e-06	2.00e-06	0.021
22400.00	4.23e-06	1.96e-06	0.031
24400.00	3.80e-06	1.92e-06	0.048
26400.00	3.33e-06	1.89e-06	0.078
28400.00	2.83e-06	1.87e-06	0.131
30400.00	2.30e-06	1.87e-06	0.218
32400.00	1.76e-06	1.89e-06	0.353
34400.00	1.21e-06	1.94e-06	0.535
36400.00	6.41e-07	2.01e-06	0.750
38400.00	7.12e-08	2.09e-06	0.973
40400.00	-4.99e-07	2.19e-06	0.820
42400.00	-1.07e-06	2.29e-06	0.642
44400.00	-1.62e-06	2.40e-06	0.498
46400.00	-2.17e-06	2.49e-06	0.384
48400.00	-2.70e-06	2.58e-06	0.295
50400.00	-3.20e-06	2.64e-06	0.225
52400.00	-3.68e-06	2.69e-06	0.170
54400.00	-4.13e-06	2.70e-06	0.126
56400.00	-4.54e-06	2.69e-06	0.091
58400.00	-4.91e-06	2.64e-06	0.063
60400.00	-5.23e-06	2.56e-06	0.041
62400.00	-5.49e-06	2.44e-06	0.024
64400.00	-5.71e-06	2.30e-06	0.013
66400.00	-5.86e-06	2.13e-06	0.006
68400.00	-5.96e-06	1.96e-06	0.002
70400.00	-6.00e-06	1.80e-06	0.001
72400.00	-5.97e-06	1.67e-06	0.000
74400.00	-5.89e-06	1.60e-06	0.000
76400.00	-5.75e-06	1.59e-06	0.000
78400.00	-5.56e-06	1.66e-06	0.001
80400.00	-5.32e-06	1.79e-06	0.003
82400.00	-5.05e-06	1.94e-06	0.009
84400.00	-4.75e-06	2.09e-06	0.023
86400.00	-4.42e-06	2.24e-06	0.048
88400.00	-4.08e-06	2.36e-06	0.083
90400.00	-3.73e-06	2.44e-06	0.126
92400.00	-3.38e-06	2.48e-06	0.172
94400.00	-3.04e-06	2.48e-06	0.220
96400.00	-2.71e-06	2.44e-06	0.267
98400.00	-2.39e-06	2.37e-06	0.313

Приложение 2. Расчеты динамики вероятности того, что типичный респондент будет чаще посещать рестораны и бары в зависимости от дохода

	Margin	Std. Err.	p-value
400.00	0.1182	0.0223	0.000
2400.00	0.1298	0.0208	0.000
4400.00	0.1416	0.0193	0.000
6400.00	0.1535	0.0178	0.000
8400.00	0.1655	0.0168	0.000
10400.00	0.1774	0.0164	0.000
12400.00	0.1891	0.0168	0.000
14400.00	0.2004	0.0180	0.000
16400.00	0.2112	0.0199	0.000
18400.00	0.2215	0.0224	0.000
20400.00	0.2312	0.0251	0.000
22400.00	0.2400	0.0280	0.000
24400.00	0.2481	0.0309	0.000
26400.00	0.2552	0.0338	0.000
28400.00	0.2614	0.0366	0.000
30400.00	0.2665	0.0394	0.000
32400.00	0.2706	0.0420	0.000
34400.00	0.2735	0.0446	0.000
36400.00	0.2754	0.0471	0.000
38400.00	0.2761	0.0496	0.000
40400.00	0.2757	0.0521	0.000
42400.00	0.2741	0.0545	0.000
44400.00	0.2714	0.0571	0.000
46400.00	0.2676	0.0596	0.000
48400.00	0.2627	0.0623	0.000
50400.00	0.2568	0.0650	0.000
52400.00	0.2499	0.0678	0.000
54400.00	0.2421	0.0706	0.001
56400.00	0.2335	0.0734	0.001
58400.00	0.2240	0.0762	0.003
60400.00	0.2139	0.0788	0.007
62400.00	0.2031	0.0812	0.012
64400.00	0.1919	0.0833	0.021
66400.00	0.1803	0.0850	0.034
68400.00	0.1685	0.0863	0.051
70400.00	0.1565	0.0870	0.072
72400.00	0.1446	0.0872	0.097
74400.00	0.1327	0.0867	0.126
76400.00	0.1210	0.0856	0.157
78400.00	0.1097	0.0839	0.191
80400.00	0.0988	0.0815	0.225
82400.00	0.0885	0.0785	0.260
84400.00	0.0787	0.0750	0.295
86400.00	0.0695	0.0711	0.328
88400.00	0.0610	0.0668	0.361
90400.00	0.0532	0.0622	0.393
92400.00	0.0460	0.0574	0.422
94400.00	0.0396	0.0526	0.451
96400.00	0.0339	0.0477	0.478
98400.00	0.0288	0.0430	0.503

Приложение 3. Код

```
/// Пункт (1,2)
//Зависимая переменная
gen visit = 1 if q46 == 2
replace visit = 2 if q46 == 3
replace visit = 3 if q46 == 1
//Регрессоры
gen male = 1 if q1==1 // Пол
replace male =0 if q1==2
gen income_ = q99/q76 // Доход на члена семьи
replace income_ = . if q99==9
gen age_ = q2 // Возраст
gen smoke_ = 1 if q34 == 1 | q34 == 2 // Курение
replace smoke_ = 0 if q34 == 3 | q34 == 4
gen health=1 if q5==1 | q5==2 // Следит ли за здоровьем
replace health=0 if q5==3 | q5==4

/// Пункт (3)
gen indicate = 1 if visit !=. & male !=. & income_ !=. & age_ !=. & smoke_ !=. & health !=.
by visit, sort: sum male income_ age_ smoke_ health if indicate==1

/// Пункт (4)
ologit visit i.male c.age c.income_ i.smoke_ i.health
est store ologit
predict pologit1 pologit2 pologit3
sum pologit1 pologit2 pologit3 visit if pologit1 !=. & visit !=.
tab visit if indicate==1

/// Пункт (5)
ologit visit i.male c.income_ c.age_ i.health
est store m0
ologit visit i.male c.age c.income_ i.health if smoke_==1
est store m_s
ologit visit i.male c.age c.income_ i.health if smoke_==0
est store m_ns
lrtest m0 (m_s m_ns)

/// Пункт (6)
ologit visit i.male c.age c.income_ i.health c.income_#c.income_ if smoke_==1
est store m_s_inc
ologit visit i.male c.age c.income_ i.health c.income_#c.income_ if smoke_==0
est store m_ns_inc

/// Пункт (7)
ologit visit i.male c.age c.income_ i.health c.income_#c.income_ if smoke_==0
est store ologit_no_smoke
ologit visit c.age c.income_ i.health c.income_#c.income_ if smoke_==0
est store ologit_no_smoke_rest
lrtest ologit_no_smoke ologit_no_smoke_rest
```

```

/// Пункт (8)
ologit visit i.male c.age c.income_ i.health c.income_#c.income_ if smoke_==0
estat ic
ologit visit c.age c.income_ i.health c.income_#c.income_ if smoke_==0
estat ic

/// Пункт (10)
tab visit if smoke_==0 & indicate==1
by visit, sort: sum income_ age_ health if indicate==1 & smoke_==0

/// Пункт (11)
ologit visit c.age c.income_ i.health c.income_#c.income_ if smoke_==0
margins, dydx(income_) predict(outcome(3)) at(health=1 income_ = (400(2000)100000))
marginsplot

/// Пункт (12)
ologit visit c.income_ c.age i.health c.income_#c.income_ if smoke_==0
margins, predict(outcome(3)) at(health=1 income_ = (400(2000)100000))
marginsplot

/// Пункт (13)
ologit visit c.income_ c.age i.health c.income_#c.income_ if smoke_==0
oparallel

/// Пункт (14)
gen income_sqr = income_^2
gologit2 visit income_ age health income_sqr if smoke_==0
est store gologit

/// Пункт (15)
gologit2 visit income_ age health income_sqr if smoke_==0, gamma
est store gologit_gamma

```