Федеральное государственное автономное образовательное учреждение высшего профессионального образования

«Национальный исследовательский университет “Высшая школа экономики”»

Санкт-Петербургский филиал федерального государственного автономного образовательного учреждения высшего профессионального образования «Национальный исследовательский университет “Высшая школа экономики”»

Образовательная программа «Экономика»

**Исследовательский проект по эконометрике**

на тему: «Структура и ценообразование на рынке наркотических веществ в интернете»

Выполнили:

Курапов Андрей (БЭК 155)  
Рычко Артем (БЭК 155)  
Слепцов Илья (БЭК 154)

Санкт-Петербург

2017

**Введение**

В современном мире вопрос частичной легализации наркотиков стоит особенно остро: многие государства значительно давно легализовали легкие наркотики (марихуана и ее образующие) и наблюдают положительную динамику оттока употребления населением алкогольных и наркотических средств; какие-то находятся на стадии принятия закона или рассмотрения инициативы внедрения, а некоторые не предпринимают никаких шагов в сторону легализации наркотиков.

Что же останавливает третьих от введения инициативы, которая положительно показывает себя в других странах? В первую очередь – незнание масштабов рынка и непонимание его структуры. Ввиду нелегального статуса торговцы наркотиками тщательно скрывают свой бизнес: продажи в основном осуществляются в темном анонимном интернете, наркодилеры и закладчики знают друг друга только по никнейму, сам товар прячется в труднодоступных и необычных местах. Выйти на след наркоторговца или хотя бы оценить их примерное количество на рынке достаточно сложно. Определение объемов теневой экономики сама по себе очень сложная задача, а если рассматривать закрытый и труднодоступный рынок наркотиков – практически нереальная.

К слову о степени вреда для организма не раз было доказано учеными меньший вред легких видов наркотиков в сравнении с алкоголем или табачными изделиям, которые до сих пор являются легальным.

С нашей точки зрения, задача определения структуры и равновесия на рынке наркотиков является актуальной, поскольку рынок является: существенным по объему, в последние годы существует тренд на легализацию некоторых видов разными правительствами, интересен экономический или поведенческо-экономический эффект существования и перехода между потребляемыми или продаваемыми товарами на рынке покупателей и продавцов, а также существование рынка как ограниченного или частично ограниченного рынка от рекламы, налогообложения и прочего.

В ходе работы проверяются следующие гипотезы:

* Состояние на рынке кроме влияния цены на количество продаваемых товаров обуславливается информацией на веб-странице продаваемого товара.
* Покупатели являются эластичными по доходу, цене и риску
* Близость рынка к монополистической конкуренции
* Возможность прибыльного и социально-значимого участия государства на рынке

В первой части исследования будут проверены первые 2 гипотезы.

**Сбор данных**

Для данного исследования был проведен «парсинг сайта» для получения данных. Парсинг – извлечение данных с сайта путем написания автоматизированного скрипта. Сайтом был выбран самый большой на данный момент незаблокированный и живой «маркетплейс» на рынке нелегальной продукции “Hydra” (http://hydra4jpwhfx4mst.onion/shops). Доступ к данному сайту осуществляется через темную часть интернета, в которую попасть можно через прокси “socks-5” или через браузер Tor, который автоматизирует этот процесс.

Для написания автоматизированного скрипта для парсинга данных использовался язык программирования и интерактивная среда разработки - python 3.6 и Jupyter notebook. Алгоритм парсинга сайта был основан на следующих этапах: получение доступа в сеть интернет с прокси “socks-5”, обход «каптчи» посредством алгоритма распознавания текста, построенный через алгоритмы машинного обучения. Каптча – процесс инициализации пользователя как человека. Следующим шагом было сбор ссылок магазинов, продающих конкретный товар, на котором было построено исследование. После был произведен парсинг информации о магазине: рейтинг, количество сделок. Затем по имеющимся ссылкам товаров был произведён автоматизированный сбор информации товара по следующим факторам: количество сделок по товару, город, количество возможных мест получения товара, количество символов в описании, комментарии (включая пользователя, время, сам комментарий) и т.д.

**Описание данных**

Наша выборка состоит из 2317 очищенных значений по наркотическому средству - марихуана, взятых с портала “Hydra”. В качестве зависимой переменной мы рассматриваем цену на коноплю в выбранном интернет-магазине данного субъекта Российской Федерации – price\_p.

Объясняющие переменные:

region – регион, к которому принадлежит магазин;

zp – средняя зарплата в регионе;

doli – доля преступлений в регионе, связанных с наркотиками, среди всех преступлений в регионе;

index – время существования интернет-магазина на портале

rating\_p – рейтинг магазина

rating\_s – рейтинг товара

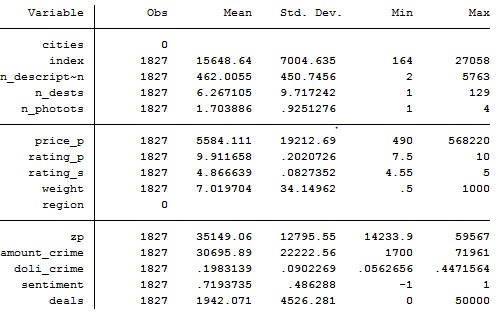
weight – вес выставленного товара

n\_photos – количество фотографий товара на сайте магазина

sentiment – сентиментальный анализ комментариев на положительные/отрицательные

n\_description – длина описания товара на сайте интернет-магазина

n\_dests – количество регионов, в которых работает данный интернет-магазин



**Спецификация модели**

Для того, чтобы грамотно подойти к выбору правильной спецификации модели необходимо ввести теорию, которая лежит в основе данного исследования. Итак, для того, чтобы найти равновесное облако цен, мы должны грамотно решить задачу как производителя, так и потребителя, путем введения в обоих случаях своих функций полезности с ограничениями. Стоит отметить, что в данной работе рассматривался только один товар, а именно «бошки»

Таким образом:

* Функция полезности покупателя с его бюджетным ограничением:

max ( – кол-во употребляемого наркотика, - кол-во потребляемых продуктов в потребительской корзине

S.t , где M, средний заработок в регионе

* Функция полезности производителя и его функция прибыли.

– функция полезности одного магазина, где заработная плата магазина в день, а – страх магазина, а именно владельцев магазина быть пойманными за продажу и хранение наркотических средств.

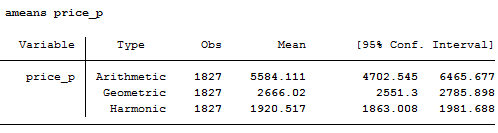
В свою очередь функция отдельно взятого магазина имеет следующий вид:

П(,q,E) = , где С – расходы магазина на заработную плату закладчиков

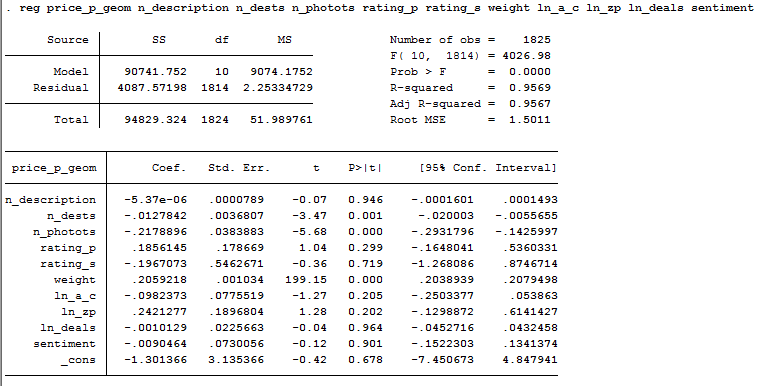
Решив задачу потребителя и производителя можно найти значения равновесных цен на рынке наркотических средств.

Выбор верной спецификации модели начался с обоснования зависимой переменной, в качестве которой выступает price\_p, а именно цена за наркотическое средство. Выбор зависимого фактора был продиктован тем, что рынок наркотических средств в России, так и в остальном мире крайне не исследован, поэтому исходя из названия исследовательского проекта было решено выявить факторы, определяющие равновесное облако цен на наркотик. В качестве объясняющий переменных нашей модели выступили как характеристика товара и магазина, а именно: вес, рейтинг товара и магазина среди покупателей, отзыв потребителей, кол-во символов в отзыве(комментарий), количество мест, где можно забрать товар, так и переменные, характеризующие социальное и экономическое положение в обществе: кол-во преступлений и заработная плата населения России по регионам.

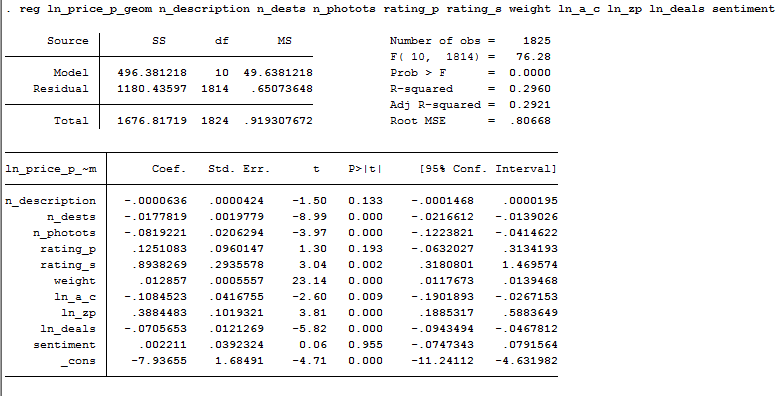
Выбирая функциональную зависимость переменной price\_p, был проведет частный тест Бокса-Кокса на выбор спецификации, а именно тест Зарембки. Было оценено две регрессии, где в первом случае в качестве зависимой переменной выступает изначальная цена, поделенная на среднее геометрическое от цены, а во втором случае под натуральным логарифмом:



**1 вариант**



**2 вариант**

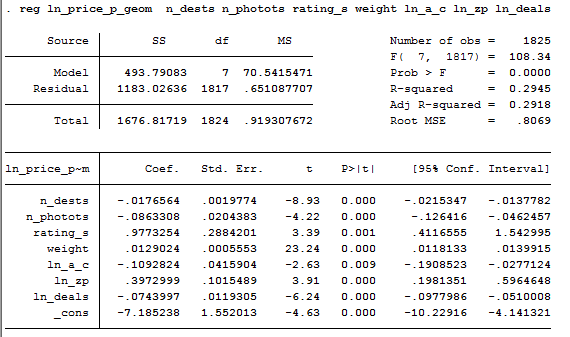


Вычисляем значение тестовой статистики:

1133.3864

Мы отчетливо видим, что значение тестовой статистики Хи квадрата намного больше критического значения, следовательно мы отвергаем нулевую гипотезу о том, что между (1) и (2) моделями нет разницы, поэтому мы выбираем модель с наименьшим RSS. Соответственно выбор падает на модель, где в качестве зависимой переменной стоит цена под натуральным логарифмом.

Двигаясь дальше в выборе наилучшей модели остановимся на аутпуте второй модели и внимательно посмотрим на значимость регрессоров, для того, чтобы откинуть незначимые. Из последней модели видно, что три регрессора стали незначимыми, а именно кол-во символов в отзыве, рейтинг продукта и оценка комментария, что в принципе ожидалось в самом начале, ведь покупателя не слишком волнует рейтинг продукта, так как, если у магазина достаточно высоких рейтинг, следовательно вероятность покупки становится больше, чем в магазине с низким рейтингом. Также оказалось вполне логичным , что кол-во символов, а именно длина комментария никак не влияет на формировании цены на наркотическое средство, ведь продавец не ссылается в ценообразовании на длину оставленных комментариев и степени их негативности. Получили следующую модель (без 3 переменных) **3 вариант**



Для того, чтобы понять, что наша модель специфицирована верно воспользуемся информационным критерием Акаике и Шварца. Посчитаем для второго и для третьего варианта значения информационного критерия.

AIC2= -.00477

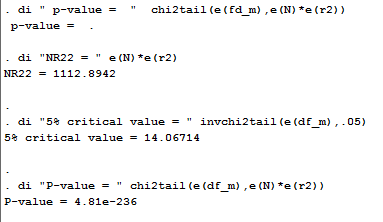
SIC2=-.0179

AIC3= -.00614

SIC3= -.0214

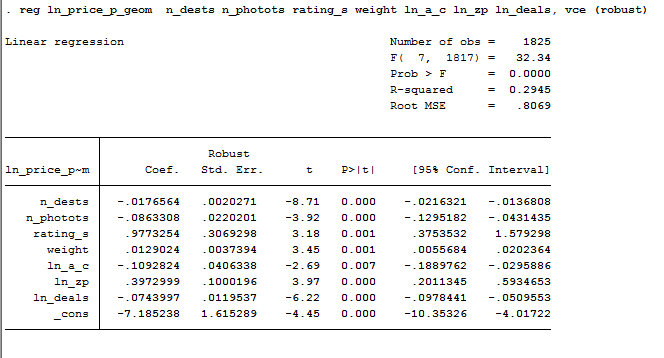
Полученные результаты свидеттельствует в пользу выбора последней модели в качестве итоговой.

Была сделана проверка на гетероскедастичность при помощи теста Вайта, получилось следующее:



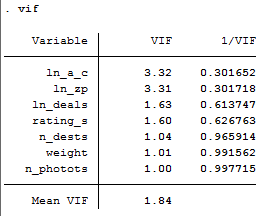
Мы видим, что нулевая гипотеза о том, что в модели есть гомоскедастичность отвергается, так как p-value меньше уровня значимости , следовательно в модели присутствует гетерскедастичность, с которой нужно бороться или хотя бы уменьшить ее влияние на нашу исходную модель.

Для того, чтобы избавиться от существующей проблемы гетероскедастичности нашей исходной модели было решено сделать робаст на ошибки. Получилось слудющее:

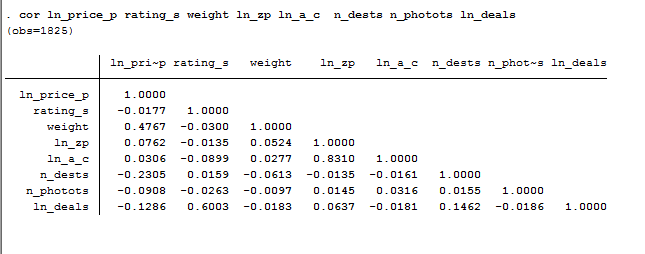


Что же изменилось, после проделанной процедуры, а именно у некоторых регрессоров изменилось как наблюдаемое значение t-статистики, так и среднеквадратическое отклонение, что дает более точную оценку доверительным интервалам, которые в большинстве случаев уменьшились, так изменилось среднеквадратическое отклоонение.

Для того, чтобы перейти к окончательной интерпретации результатов был проведено исследование на существование мультиколлинеарности путем отыскания коэффициента вздутия дисперсии и постройкой корреляционной матрицы.

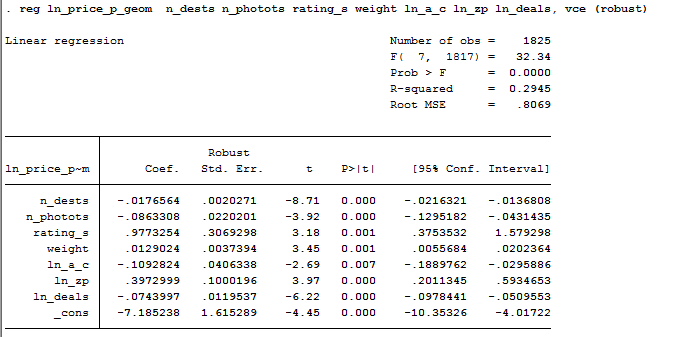


Из таблицы видно, что в модели отсутствует мультиколленеарность , так коэффициент вздутия дисперсии слишком маленький. Об этом свидетельствует еще корреляционная матрица.



Из корреляционной таблицы видно, что умеренную корреляцию с логарифмом цены наркотика имеет вес самого наркотика, что обосновывается довольно просто, чем больше вес покупаемого наркотического средства, тем выше цена за наркотический препарат, ведь зависимость положительная и прямо пропорциональная. Также следует отметить, что кол-во точек, где можно забрать товар влияет умеренно, обратно пропорционально на логарифм цены наркотика. Из матрицы также очевидно выявлено, что такой регрессор как кол-во преступлений, связанными с наркотическими средствами, имеют очень сильную положительную корреляцию с заработной платой людей, что обосновывается с точки зрения человеческих предпочтений, чем больше человек зарабатывает, тем больше ему хочется попробовать в жизни, наркотик является выходом из этой ситуации.

В итоге было решено остановиться на следующей модели:



Из полученных оцененных коэффициентов видно, что все регрессоры имеют ожидаемые знаки, которые предполагались изначально. Например, с возрастанием кол-ва обнаруженных преступлений (ln\_a\_c) цена на накротическое средство падает, что является очень логичным, так как с возрастанием эффективности органов правопорядка , страх быть пойманным за употребление наркотических средств возрастает, следовательно люди не покупают наркотик, что ведет к снижению цены на наркотический препарат

Интерпретация:

С увеличением кол-ва точек распространения на 1 точку приведет к снижению цены на наркотическое средство на 1,7%. Это связано с тем, что увеличение точек ведет к увеличению слежки за доставщиками наркотических средств, а также самими покупателями, что ведет к снижению спроса. Тогда магазин вынужден будет снизить цену на свой товар

С увеличением кол-ва фотографий продукта (бошки) на 1 приведет к снижению цены на 8,6%. Предположение о том, что покупатель будет всматриваться в выставленный товар и он ему не будет нравится. То есть нахождение изъянов в выращенном товаре.

:С увеличением рейтинга магазина на 1 пункт, ведет к увеличению цены на 97%, что совершенно оправданно, ведь чтобы стать популярным магазином на платформе, не имея никаких покупателей и рейтнга, стоит снизить цену до своего минимума. Так наработается необходимая клиентская база, в последствии которая распространит информацию о магазине среди других людей, возрастет оборот магазина, следовательно возрастет рейтинг. Магазины с наибольшим рейтингом всплывают на ресурсе в поисковом запросе самые первые, в отличие от магазинов с низким рейтингом.

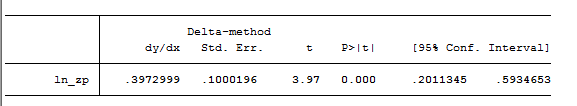
: с увеличением веса на 1 г, ведет к увеличению цены на 1,2%.

этосвязано с тем, что страх быть пойманным за употребление наркотических средств возрастает, следовательно люди не покупают наркотик, что ведет к снижению цены на наркотический препарат

С увеличением заработной платы населения на 1%, цена на наркотик увеличивается на 0,397%

С увеличением кол-ва сделок на 1% ведет к уменьшению цены на 1 процент, связано с отповизной приобретаемой цены.

Отвечая на одну из поставленных гипотез, мы можем найти эластичность цены на наркотик по доходу, а именно заработной плате и прояснить, чем она является



Мы видим, что эластичность цены по доходу равна 0,397, что с увеличением заработной платы населения на 1%, цена на наркотик увеличивается на 0,397%.

**Выводы и перспективы**

В первой части исследования был проведен анализ структуры рынка и ценообразования на рынке наркотических средств на самом большом незаблокированным маркетпэйсе России. Были выявлены факторы, влияющие на ценообразование на данном рынке такие как: количество совершенных сделок; факторы, описывающие качество товара - комментарии, рейтинг товара и магазина, длина описания и количество возможных мест сбыта товара; вес товара; уровень зарплат в регионе и количество преступлений по законам, которые напрямую повышают риск совершения купли-продажи на рынке. Была доказана положительная связь уровня зарплат, количества преступлений и величины продаваемого веса. Также была полученная убывающая отдача цены от продаваемого веса, что является «силой оптовизны» товара.

Задачами второй части исследования являются сбор и анализ панельных данных по всем видам наркотических средств на данным маркетплейсе. Это позволит выявить полную динамику изменения цен от влияющих на нее факторов и установить причинность выхода и входа игроков на рынке, что влияет на товарооборот, построение матрицы эластичностей субститутов и определение силы экстерналий, асимметрии информации между игроками рынка. Следующей задачей является определения оптимального налогообложения рынка наркотических средств, которое позволит сократить количество людей, употребляющих наркотики, или найти оптимальный уровень налогообложения, приносящий максимальный налоговый сбор государству. Важным исследовательским вопросом во второй части исследования является причины действий игроков на изменения разных факторов в динамике. Например, увеличение активности правоохранительных органов или изменение тональности комментариев определенных пользователей (нахождение opinion-maker-ов на рынке). В долгосрочной динамике можно будет наблюдать причины дифференциации товаров и способы рекламы в ограниченной среде.