

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)

## **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.

Тем временем, в мире развивается большое количество дискуссий о целесообразности не просто легализации наркотиков, а полной замены алкогольных средств на наркотические. Учеными доказан как меньший вред на человеческий организм наркотиков в сравнении с алкоголем, так и более яркий эффект от употребления наркотиков.

С нашей точки зрения, задача поиска структуры и равновесия на рынке наркотиков является актуальной, но в то же время трудной и очень интересной. Мы проанализировали стоимость и объем наркотических товаров на крупнейшем российском агрегаторе интернет-магазинов по продаже наркотиков – hydra и выявили основные детерминанты образования цен на наркотические средства.

## **Извлечение 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://hydra4jpwfhx4mst.onion/shops>). Доступ к данному сайту осуществляется через темную часть интернета, в которую попасть можно через прокис “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 «капчи» посредством алгоритма распознавания текста построенный через алгоритмы машинного обучения. Капча – процесс инициализации пользователя как человека, а не автоматизированного парсера. 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 компонентам: количество сделок по товару, город, количество возможных мест получения товара, количество символов в описании, комментарии (включая пользователя, время, сам комментарий) и т.д. (можно увидеть полный список в следующем разделе).

## Описание переменных

Наша выборка состоит из 2317 очищенных значений по конопле, взятых с портала hydra (<http://hydraruzxpnew4af.onion/>). 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 – логарифм средней зарплаты 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 – эмоциональный анализ комментариев на положительные/отрицательные

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

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

Variable	Obs	Mean	Std. Dev.	Min	Max
cities	0				
index	1827	15648.64	7004.635	164	27058
n_description	1827	462.0055	450.7456	2	5763
n_dests	1827	6.267105	9.717242	1	129
n_photos	1827	1.703886	.9251276	1	4
price_p	1827	5584.111	19212.69	490	568220
rating_p	1827	9.911658	.2020726	7.5	10
rating_s	1827	4.866639	.0827352	4.55	5
weight	1827	7.019704	34.14962	.5	1000
region	0				
zp	1827	35149.06	12795.55	14233.9	59567
amount_crime	1827	30695.89	22222.56	1700	71961
doli_crime	1827	.1983139	.0902269	.0562656	.4471564
sentiment	1827	.7193735	.486288	-1	1
deals	1827	1942.071	4526.281	0	50000

## Предполагаемая спецификация модели и ее обоснование

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:  
 $U(x_1, Y_i) \longrightarrow \max (x_1, Y_i \geq 0)$ ,  $x_1$  – кол-во употребляемого наркотика,  $Y_i$  - кол-во потребляемых продуктов в потребительской корзине

S.t  $x_1 p_1 + p_2 Y_1 + \dots + p_n Y_{n-1} + n = M$ , где  $M$ , средний заработок в регионе

- Функция полезности производителя и его функция прибыли.

$U\left(\frac{\pi Q}{N}, \frac{E}{N}\right)$  – функция полезности одного магазина, где  $\frac{\pi Q}{N}$  заработная плата магазина в день, а  $\frac{E}{N}$  – страх магазина, а именно владельцев магазина быть пойманными за продажу и хранение наркотических средств.

В свою очередь функция отдельно взятого магазина имеет следующий вид:

$\Pi(\pi, q, E) = \pi * q_{max} - C - \frac{E}{N}$ , где  $C$  – 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 проведен 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 логорифмом:

`ameans price_p`

Variable	Type	Obs	Mean	[95% Conf. Interval]	
price_p	Arithmetic	1827	5584.111	4702.545	6465.677
	Geometric	1827	2666.02	2551.3	2785.898
	Harmonic	1827	1920.517	1863.008	1981.688

**1 вариант**

. reg price_p_geom n_description n_dests n_photots rating_p rating_s weight ln_a_c ln_zp ln_deals sentiment						
Source	SS	df	MS	Number of obs = 1825		
Model	90741.752	10	9074.1752	F( 10, 1814) = 4026.98		
Residual	4087.57198	1814	2.25334729	Prob > F = 0.0000		
				R-squared = 0.9569		
				Adj R-squared = 0.9567		
Total	94829.324	1824	51.989761	Root MSE = 1.5011		

  

price_p_geom	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
n_description	-5.37e-06	.0000789	-0.07	0.946	-.0001601	.0001493
n_dests	-.0127842	.0036807	-3.47	0.001	-.020003	-.0055655
n_photots	-.2178896	.0383883	-5.68	0.000	-.2931796	-.1425997
rating_p	.1856145	.178669	1.04	0.299	-.1648041	.5360331
rating_s	-.1967073	.5462671	-0.36	0.719	-1.268086	.8746714
weight	.2059218	.001034	199.15	0.000	.2038939	.2079498
ln_a_c	-.0982373	.0775519	-1.27	0.205	-.2503377	.053863
ln_zp	.2421277	.1896804	1.28	0.202	-.1298872	.6141427
ln_deals	-.0010129	.0225663	-0.04	0.964	-.0452716	.0432458
sentiment	-.0090464	.0730056	-0.12	0.901	-.1522303	.1341374
_cons	-1.301366	3.135366	-0.42	0.678	-7.450673	4.847941

## 2 вариант

. reg ln_price_p_geom n_description n_dests n_photots rating_p rating_s weight ln_a_c ln_zp ln_deals sentiment						
Source	SS	df	MS	Number of obs = 1825		
Model	496.381218	10	49.6381218	F( 10, 1814) = 76.28		
Residual	1180.43597	1814	.65073648	Prob > F = 0.0000		
				R-squared = 0.2960		
				Adj R-squared = 0.2921		
Total	1676.81719	1824	.919307672	Root MSE = .80668		

  

ln_price_p~m	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
n_description	-.0000636	.0000424	-1.50	0.133	-.0001468	.0000195
n_dests	-.0177819	.0019779	-8.99	0.000	-.0216612	-.0139026
n_photots	-.0819221	.0206294	-3.97	0.000	-.1223821	-.0414622
rating_p	.1251083	.0960147	1.30	0.193	-.0632027	.3134193
rating_s	.8938269	.2935578	3.04	0.002	.3180801	1.469574
weight	.012857	.0005557	23.14	0.000	.0117673	.0139468
ln_a_c	-.1084523	.0416755	-2.60	0.009	-.1901893	-.0267153
ln_zp	.3884483	.1019321	3.81	0.000	.1885317	.5883649
ln_deals	-.0705653	.0121269	-5.82	0.000	-.0943494	-.0467812
sentiment	.002211	.0392324	0.06	0.955	-.0747343	.0791564
_cons	-7.93655	1.68491	-4.71	0.000	-11.24112	-4.631982

Вычисляем значение тестовой статистики:

$$\chi^2 = \frac{n}{2} \left| \ln \left( \frac{RSS_1}{RSS_2} \right) \right|$$

$$\chi^2 = 1133.3864$$

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**

. reg ln_price_p_geom n_dests n_photots rating_s weight ln_a_c ln_zp ln_deals						
Source	SS	df	MS	Number of obs = 1825		
Model	493.79083	7	70.5415471	F( 7, 1817) = 108.34		
Residual	1183.02636	1817	.651087707	Prob > F = 0.0000		
				R-squared = 0.2945		
				Adj R-squared = 0.2918		
				Root MSE = .8069		
Total	1676.81719	1824	.919307672			

  

ln_price_p~m	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
n_dests	-.0176564	.0019774	-8.93	0.000	-.0215347	-.0137782
n_photots	-.0863308	.0204383	-4.22	0.000	-.126416	-.0462457
rating_s	.9773254	.2884201	3.39	0.001	.4116555	1.542995
weight	.0129024	.0005553	23.24	0.000	.0118133	.0139915
ln_a_c	-.1092824	.0415904	-2.63	0.009	-.1908523	-.0277124
ln_zp	.3972999	.1015489	3.91	0.000	.1981351	.5964648
ln_deals	-.0743997	.0119305	-6.24	0.000	-.0977986	-.0510008
_cons	-7.185238	1.552013	-4.63	0.000	-10.22916	-4.141321

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 свидетельствует в пользу выбора последней модели в качестве итоговой.

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

```
. di " p-value = " chi2tail(e(df_m),e(N)*e(r2))
p-value = .

. di "NR22 = " e(N)*e(r2)
NR22 = 1112.8942

.
. di "5% critical value = " invchi2tail(e(df_m),.05)
5% critical value = 14.06714

.
. di "P-value = " chi2tail(e(df_m),e(N)*e(r2))
P-value = 4.81e-236
```

Мы видим, что нулевая гипотеза о том, что в модели есть гомоскедастичность отвергается, так как p-value меньше уровня значимости, следовательно в модели присутствует гетероскедастичность, 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 следующее:

```
. reg ln_price_p geom n_dests n_photots rating_s weight ln_a_c ln_zp ln_deals, vce (robust)
```

Linear regression

Number of obs = 1825  
F( 7, 1817) = 32.34  
Prob > F = 0.0000  
R-squared = 0.2945  
Root MSE = .8069

ln_price_p~m	Robust		t	P> t	[95% Conf. Interval]	
	Coef.	Std. Err.				
n_dests	-.0176564	.0020271	-8.71	0.000	-.0216321	-.0136808
n_photots	-.0863308	.0220201	-3.92	0.000	-.1295182	-.0431435
rating_s	.9773254	.3069298	3.18	0.001	.3753532	1.579298
weight	.0129024	.0037394	3.45	0.001	.0055684	.0202364
ln_a_c	-.1092824	.0406338	-2.69	0.007	-.1889762	-.0295886
ln_zp	.3972999	.1000196	3.97	0.000	.2011345	.5934653
ln_deals	-.0743997	.0119537	-6.22	0.000	-.0978441	-.0509553
_cons	-7.185238	1.615289	-4.45	0.000	-10.35326	-4.01722

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

```
. vif
```

Variable	VIF	1/VIF
ln_a_c	3.32	0.301652
ln_zp	3.31	0.301718
ln_deals	1.63	0.613747
rating_s	1.60	0.626763
n_dests	1.04	0.965914
weight	1.01	0.991562
n_photots	1.00	0.997715
Mean VIF	1.84	

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

```
. cor ln_price_p rating_s weight ln_zp ln_a_c n_dests n_photots ln_deals  
(obs=1825)
```

	ln_pri~p	rating_s	weight	ln_zp	ln_a_c	n_dests	n_phot~s	ln_deals
ln_price_p	1.0000							
rating_s	-0.0177	1.0000						
weight	0.4767	-0.0300	1.0000					
ln_zp	0.0762	-0.0135	0.0524	1.0000				
ln_a_c	0.0306	-0.0899	0.0277	0.8310	1.0000			
n_dests	-0.2305	0.0159	-0.0613	-0.0135	-0.0161	1.0000		
n_photots	-0.0908	-0.0263	-0.0097	0.0145	0.0316	0.0155	1.0000	
ln_deals	-0.1286	0.6003	-0.0183	0.0637	-0.0181	0.1462	-0.0186	1.0000

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:

```
. reg ln_price_p_geom n_dests n_photots rating_s weight ln_a_c ln_zp ln_deals, vce (robust)
```

```
Linear regression                               Number of obs =   1825
                                                F(   7,  1817) =   32.34
                                                Prob > F       =   0.0000
                                                R-squared      =   0.2945
                                                Root MSE     =   .8069
```

ln_price_p~m	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
n_dests	-.0176564	.0020271	-8.71	0.000	-.0216321	-.0136808
n_photots	-.0863308	.0220201	-3.92	0.000	-.1295182	-.0431435
rating_s	.9773254	.3069298	3.18	0.001	.3753532	1.579298
weight	.0129024	.0037394	3.45	0.001	.0055684	.0202364
ln_a_c	-.1092824	.0406338	-2.69	0.007	-.1889762	-.0295886
ln_zp	.3972999	.1000196	3.97	0.000	.2011345	.5934653
ln_deals	-.0743997	.0119537	-6.22	0.000	-.0978441	-.0509553
_cons	-7.185238	1.615289	-4.45	0.000	-10.35326	-4.01722

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 наркотическое 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:

$V_1$ : 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

$V_2$ : 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.

$V_3$ : 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 рейтинга, стоит снизить цену до своего минимума. Так нарабатывается 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.

$V_4$ : с увеличением веса на 1 г, ведет к увеличению цены на 1,2%.

$V_5$ : С увеличением кол –

ва пойманных людей, связанных с наркобизнесом на 1 %, цена понижается на 0,0109% это связано с тем, что страх быть пойманным за употребление наркотических средств возрастает, следовательно люди не покупают наркотик, что ведет к снижению цены на наркотический препарат

$V_6$ : С увеличением заработной платы населения на 1%, цена на наркотик увеличивается на 0,397%

$V_7$ : 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

	Delta-method		t	P> t	[95% Conf. Interval]	
	dy/dx	Std. Err.				
ln_zp	.3972999	.1000196	3.97	0.000	.2011345	.5934653

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%.

### Заключение и перспективы развития

В данной работе было проведено исследование о структуре рынка и ценообразовании на рынке наркотических средств в интернете. 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. Была продемонстрирована рентабельность входа государства на рынок и положительное влияние параметров веб-страницы продаваемого товара на его цену.

Перспективой развития данной работы является сбор панельных данных по всем видам наркотических средств на данном маркетплейсе, а также в маркетплейсе других стран. Это позволит выявить полную динамику изменения цен от влияющих на нее факторов и понять из-за чего уходят игроки (покупатели или продавцы) с рынка, что влияет на количество покупаемых товаров, построение матрицы эластичности субститутов и определение силы экстерналий, асимметрии информации между игроками рынка. Также будет возможно установление налогообложения рынка наркотических средств, которое позволит сократить количество людей, употребляющие наркотики, или найти оптимальный уровень налогообложения, приносящий максимальный налоговый сбор государству. Важной перспективой исследования является причина реакции игроков на изменения разных факторов. Например увеличение активности правоохранительных органов или сила комментариев определенных пользователей (нахождение opinion-maker-ов). В динамике года можно будет наблюдать причины дифференциации товаров, на чем можно наблюдать переход рынка с преобладанием рынком характеристик рынка монополистической конкуренции в рынок близкой к совершенной конкуренции, благодаря вариации факторов.