

ПРОЕКТИРОВАНИЕ МОДЕЛИ УРОВНЯ ПРЕСТУПНОСТИ В РОССИИ

Матлашевская В.В.

Национальный Исследовательский Университет «Высшая Школа Экономики» (614070, г. Пермь, ул. Студенческая, д. 38), e-mail: v.matlashevskaja@yandex.ru

В работе представлено исследование по моделированию регионального уровня преступности в Российской Федерации, состоящее из теоретического бэкграунда и практической реализации. Теоретическая часть включает в себя теоретический обзор существующей литературы зарубежных и отечественных ученых за последние двадцать лет и математических методов, описывающих как простейшие эконометрические принципы, так и более сложные этапы выбора моделей. В исследовании также отражен процесс проведения эконометрического моделирования от экспорта данных до сравнения построенных моделей и интерпретации итогов, результатом которого является модель уравнения с фиксированными эффектами зависимости регионального уровня преступности (суммарный показатель, включающий убийство, умышленное причинение вреда здоровью, изнасилование, кража, грабёж, разбой, вымогательство, хулиганство) от различных факторов (численность постоянного населения, число прерываний беременности или абортов, количество впервые ВИЧ-инфицированных, численность воспитанников школ-интернатов, численность детей, не обучающихся в образовательных организациях, количество проданной алкогольной продукции, миграционный прирост, количество впервые больных с диагнозом «наркомания», численность выпускников образовательных учреждений, не получивших аттестат). Полученное уравнение является основой разработанного приложения посредством языка программирования C# на Windows Forms для прогнозирования значения уровня преступности в отдельных регионах Российской Федерации.

Ключевые слова: уровень преступности, эконометрическое моделирование, анализ панельных данных, разработка приложения.

CRIME RATE MODEL DESIGN FOR RUSSIA

Matlashevskaja V.V.

The work presents the theoretical background and the practical implementation. The theoretical part includes the literature review of the existing foreign and domestic scientists works over the past twenty years and mathematical methods that describe both the simplest econometric principles and more complex lines of the models selecting. The study also reflects the process of carrying out econometric modeling from data export to comparison of constructed models and the result interpretation, outcoming with the fixed effects model equation of regional crime rate (total value including murder, intended bodily harm, rape, theft, brigandry, robbery, extortion, hooliganism) from various factors (the number of permanent residents, the number of abortions, the number of first-time HIV-infected, the number of pupils in boarding schools, the number of children not studying in educational organizations, the number of sold alcohol products, the net migration, the number of first-time patients diagnosed with "drug addiction", the number of graduates educational institutions that did not receive a certificate). The final equation is the basis of the developed application set up by the C # programming language on Windows Forms in order to predict the crime level in certain regions of the Russian Federation.

The Key Words: crime rate, econometric modeling, panel data analysis, application development.

Проблема борьбы с преступностью является одним из самых важных вопросов любого общества. Преступность выступает зеркальным отражением состояния государства, тенденция динамики которого влияет на развитие общества.

Вопрос об уровне преступности и возможности его изучения всегда остро поднимается как в национальных исследованиях, так и в международных. Однако, пункт о выявлении типа зависимости, факторов, их анализ и интерпретация остается малоизученным. На сегодняшний

день произведены исследования, которые затрагивают либо эконометрическую составляющего вопроса, либо техническую, но существует недостаток работ, объединяющих и разработку модели зависимости, и проектирование приложения, в области прогнозирования уровня преступности по регионам России.

Так, современные средства автоматизации процессов охватывают различные предметные области от бизнеса до политики, от бытовых функций до международной экономики. Однако, существует пробел в области факторов, влияющих на региональный уровень преступности в РФ.

Актуальность работы заключается в остроте поднимаемого вопроса; по данным МВД РФ количество совершенных преступлений за 2017 год составило более 2 миллионов единиц, а за первый квартал 2018 года более 640 тысяч преступлений – так, государство находится в поиске новых рычагов снижения показателя уровня преступности.

Кроме того, на российском рынке отсутствует адаптированное под широкий круг пользователей приложение, с автоматизированным набором методов, для прогнозирования уровня преступности на основе эконометрических моделей по ключевым параметрам – социального, экономического и демографического характера.

Вопрос о регулировании уровня преступности всегда занимал особое место в сознаниях людей, и многие авторы поднимали в своих работах вопрос о выявлении влияющих факторов на уровень преступности.

Основываясь на исследовании экономиста Стивена Левитта [6], в котором была выявлена и обоснована зависимость уровня преступности в США в 90-е годы от таких факторов, как количество аборт (с соответствующей разницей в годах), ситуацией на рынке наркотиков и огнестрельного оружия, регулирование закона о высшей мере наказания; стоит заметить «нестандартность» выбранных факторов, поскольку большей популярностью в исследованиях пользуются такие социальные и экономические регрессоры, как уровень заработной платы, ВВП, количество безработных и другие. Взяв во внимание пример американских ученых, мною будет проделана аналогичная работа на российской аудитории. Однако, несмотря на бóльшую изученность в данной области зарубежных специалистов, информационное приложение до настоящего момента остается не разработанным для широкого круга пользователей – обычных граждан, специальных служб, государственных органов.

Кроме исследования Левитта ряд ученых также затрагивал данный вопрос. Кристиан Поп-Элечес [9] отразил ситуацию в Румынии в 1970-х годах, когда Николае Чаушеску стал президентом Социалистической Республики Румынии. В течение своего правления он провел

ряд изменений для страны, одним из которых стал запрет аборт по всей стране; что в результате привело к увеличению уровня преступности на 30% уже к 1990-м годам.

Также, мировому сообществу были представлены такие работы как «Crack and Homicide in New York City: A Case Study in the Epidemiology of Violence» (Paul J. Goldstein, Patricia A. Belucci) [8], которая отражает причину 25% убийств, связанных с наркотиками; «More Guns, More Crime» (Mark Duggan) [7], – зависимость огнестрельного оружия и уровня преступности.

Касательно отечественных ученых, такие влиятельные экономисты как, Андриенко Ю.В [1]. занимался изучением преступного поведения в России; в его исследовании были изучены такие регрессоры, как: доля населения, злоупотребляющего алкоголем и наркотиками, средний уровень образования, реальный доход, коэффициент Джини (мера неравенства в доходах), уровень безработицы и вектор других индикаторов.

Основываясь на проведенном исследовании литературы, были выбраны и экспортированы[4] следующие переменные: в качестве зависимой переменной была взята сумма показателей: убийство (по ст.105 УК РФ), умышленное причинение вреда здоровью (по ст.111 УК РФ), изнасилование (по ст.131 УК РФ), кража (по ст.158 УК РФ), грабёж (по ст.161 УК РФ), разбой (по ст.162 УК РФ), вымогательство (по ст.163 УК РФ), хулиганство (по ст.213 УК РФ), выраженная в единицах совершенных преступлений; в качестве объясняющих переменных показатели: численность постоянного населения (чел.), число прерываний беременности (ед.), количество впервые ВИЧ-инфицированных (чел.), численность воспитанников школ-интернатов (чел.), численность детей, не обучающихся в образовательных организациях (чел.), количество проданной алкогольной продукции (тыс. дал.), миграционный прирост (чел.), количество впервые больных с диагнозом «наркомания» (чел.), численность обучающихся, не получивших аттестат (чел.). В работе используются данные за 3 года с 2011 по 2013 года; такой период выбора данных обусловлен полнотой информации по различным факторам, совпадающих во временных отрезках (источник данных – ЕМИСС Государственная Статистика - fedstat.ru).

Поскольку данные имеют панельный тип, то есть могут быть оценены в трех измерениях: год, субъект и показатель; следующим шагом после выгрузки данных будет проведение разведочного анализа, который включает в себя получения описательной статистики, построения корреляционной матрицы, масштабирование переменных. Описательная статистика данных показывает основные математические характеристики выбранного набора данных от простейших показателей минимума и максимума выборки, до показателей медианы, дисперсии, коэффициентов эксцесса и асимметрии. При помощи

описательной статистики данных были получены сведения о показателях в разрезе годов, например таблица 1 отображает описательную статистику данных в 2013 году.

Таблица 2.4. Описательная статистика данных 2013 года

Показатель	Среднее	Медиана	Мода	Стандартное отклонение	Дисперсия	Экссесс	Асимметричность	Минимум	Максимум
Переменная									
<i>crime</i>	13148,54	8178	Н/Д	15045,28	226360510,3	18,05	3,44	225	107072
<i>popul</i>	1755195,17	1197040	Н/Д	1754055,7	3,08E+12	14,29	3,11	42907	12043893
<i>abort</i>	10860,66	8301	Н/Д	8611,2	74152820,79	1,34	1,33	471	40217
<i>hiv</i>	44,42	29	18,3	41,85	1751,49	3,57	1,78	3	217
<i>c_home</i>	189,36	148	0	218,61	47790,16	12,39	2,79	0	1439
<i>no_edu</i>	316,76	171	280	333,15	110987,43	2,28	1,69	5	1352
<i>alco</i>	236,97	69,06	Н/Д	735,84	541459,34	57,73	7,2	0	6276,36
<i>net_migr</i>	3660,13	2107	Н/Д	5180,91	26841865,24	20,9	3,86	-337	37039
<i>drug</i>	10,28	8	4,2	7,19	51,67	1,07	1,09	0	36

где *crime* – это суммарный показатель уровня преступности; *popul* – количество населения; *abort* – количество совершенных аборт; *hiv* – количество впервые ВИЧ-инфицированных; *c_home* – численность воспитанников школ-интернатов; *no_edu* – количество детей, не обучающихся в образовательных организациях; *alco* – количество проданного алкоголя; *net_migr* – миграционный прирост, *drug* – количество пациентов с диагнозом «наркомания» впервые.

Описательный анализ включает расчеты медианы, моды, стандартного отклонения, дисперсии, коэффициентов эксцесса и асимметрии, минимума и максимума. Медиана – значение, которое разбивает выборку на равные части; ее отличие от среднего значения состоит в том, что медиана – это реальное значения из выборки, которое занимает «серединную» позицию в упорядоченной выборке. Мода – значение, которое наиболее часто встречается в выборке. При анализе данных по 83 регионам достаточно редко будут попадать повторяющиеся значения, поэтому данный показатель может принимать нулевое значение – те регионы, по которым не было информации, и принимать «0» за наиболее часто встречающееся значение. Рассеивание величины показывает стандартное отклонение в сравнении с ее математическим ожиданием. Стандартное отклонение несет в себе математический смысл для разведочного анализа. Дисперсия это стандартное отклонение в квадрате, показывающая на сколько единиц от среднего значения могут отличаться данные. Коэффициент эксцесса – это показатель остроты распределения величины в выборке. При нормальном распределении коэффициент стремится к нулю, при «приплюснутой» вершине – меньше нуля, а при «острой» - больше нуля. Коэффициент асимметрии указывает на «сдвиг» вершины: если коэффициент больше нуля, то правый хвост распределения длиннее левого, а

если отрицательный, то левый хвост более длинный. Минимум и максимум отражают наибольшее и наименьшее значение по выборке.

Анализируя таблицу 1, можно сделать заключение, что в 2013 году уровень преступности (*crime*) находился в пределах от 225 до 107072 единицы преступления, с медианой в 8178 единицы, имея асимметричное распределение. Максимальное значение пациентов с диагнозом «наркомания», поставленным впервые (*drug*), составило 36 человек.

В статистике корреляционный анализ применяется при обработке статистических данных, с целью изучения взаимосвязей между всеми переменными выборки. Коэффициент принимает значения $[1; 1]$, при стремлении значения коэффициента корреляции к единице – можно заявлять о наличии прямой зависимости переменных друг от друга (часто при попадании величины в промежуток $[0,75; 1]$, свидетельствуется о наличие зависимости переменных); о наличии обратной зависимости будет говорить значения коэффициента из промежутка $[-1; -0,75]$; если же коэффициент корреляции стремится к нулю, то можно говорить об отсутствии зависимости[5]. Так, в таблице 2 приведены коэффициенты корреляции.

Таблица 2.5. Корреляционная матрица

	<i>popul</i>	<i>abort</i>	<i>hiv</i>	<i>c home</i>	<i>no edu</i>	<i>alco</i>	<i>net migr</i>	<i>drug</i>	<i>no dipl</i>
<i>popul</i>	1								
<i>abort</i>	0,77	1							
<i>hiv</i>	0,35	0,6	1						
<i>c home</i>	0,27	0,38	0,38	1					
<i>no edu</i>	0,3	0,51	0,3	0,3	1				
<i>alco</i>	0,69	0,28	0,09	0,11	-0,01	1			
<i>net migr</i>	0,68	0,6	0,32	0,16	0,13	0,3	1		
<i>drug</i>	0,24	0,39	0,43	0,29	0,23	0,05	0,16	1	
<i>no dipl</i>	0,34	0,3	0,13	0,08	0,29	0,04	0,19	0,13	1

По данной таблице видно, что среди переменных присутствует преимущественно прямая зависимость, в отдельных ситуациях коэффициент корреляции находится в пределах $[0,75; 1]$; это пример коэффициента между показателем численности населения и количеством совершенных аборт (коэффициент корреляции равен 0,77). Это обусловлено тем, что в регионах с большим количеством людей, больше вероятность совершения прерывания беременности. Данный высокий коэффициент корреляции не является критическим в такой интерпретации.

Следующий этап – масштабирование переменных или взятие логарифма от числа переменной. Логарифмирование переменных можно интерпретировать по-разному в зависимости от набора интерпретируемых переменных, например если берутся логарифмы зависимой и объясняющей переменных, то коэффициент при факторной переменной выступает в роли эластичности, показывающий на сколько процентов изменится первая переменная при изменении второй на 1%, соответственно. Если логарифмированию

подвергается только эндогенная переменная, а экзогенная при этом остается в первоначальном виде, то параметр при объясняющем факторе будет говорить, что при изменении X (экзогенная переменная) на 1 единицу, Y (эндогенная переменная) изменится на 1%. В обратном случае (первоначальное значение Y и логарифмированное значение X) коэффициент (b_i) при факторе будет свидетельствовать, что при изменении X на 1%, Y изменится на $\frac{b_i}{100}$ единиц.

Методом пошаговой регрессии был выбран следующий набор переменных, который отвечает в модели типа *pooled* за наиболее точные показатели: в качестве эндогенных: l_crime (логарифм от *crime*), в качестве экзогенных: *abort*, *c_home*, *alco*, l_hiv (логарифм от *hiv*), l_net_migr (логарифм от *net_migr*), l_drug (логарифм от *drug*), l_no_dipl (логарифм от *no_dipl*).

Используя полученный набор переменных, были также построены эконометрические модели типа *fixed effects model* и *random effects model*; среди которых была выбрана наиболее точная методом проведения попарного сравнения этих моделей (критерий Хаусмана, критерий Бройша-Пагана, тест на различие констант в группах[2]). Наиболее точной была выбрана модель с фиксированными эффектами, которая имеет вид:

$$\widehat{l_crime}_i = 8,48 + 0,00003 * abort_i + 0,0002 * c_home_i - 0,000001 * alco_i - 0,07 * l_hiv_i - 0,003 * l_net_migr_i + 0,08 * l_drug_i + 0,05 * l_no_dipl_i$$

Модель с фиксированными эффектами показывает следующие переменные с 1% уровнем значимости: *const*, *abort*, l_drug ; с 5% - l_no_dipl ; с 10% - l_hiv .

LSDV (*least square dummy variable*) R-square, как отмечалось ранее, это R-square для LSDV модели, которая учитывает сам тип панельных данных. *Within R-square* (или «в пределах R-square») – является скорректированным R-square для модели с фиксированными эффектами. Здесь, 99% и 35% коэффициенты детерминации говорят о разности оценивания значимости объяснения внутригрупповой и межгрупповой вариации в показателях. При оценке модели с фиксированными эффектами выбирается внутригрупповой R-square или «в пределах R-square», так как он наиболее широко объясняет внутри групповую вариацию, снижая сам показатель в сравнении с LSDV R-square, который соответствует стандартному определению оценки параметров модели с помощью МНК.

Проверив полученную модель на наличие гетероскедастичности тестом «*groupwise*» в *Gretl*, были получены результаты о ее наличии. В нашей модели имеются гетерогенные наблюдения, такой вид «нарушения» может быть объяснен характерными изменениями в данных, которые описываются в один момент времени но по различным субъектам, регионам, фирмам, группам и т.д. Неоднородность дисперсий ошибок возникает как результат структурных изменений в экономике, социальной экономике, экономике географии[3]. Таким

образом, в нашей модели наличие гетероскедастичности не является критическим нарушением стандартных предположений.

Интерпретация полученной модели состоит в объяснении ее коэффициентов. Так, увеличение уровня аборт на 1 единицу приведет к увеличению уровня преступности на 0,003%; увеличение количества воспитанников школ-интернатов на 1 единицу – увеличение на 0,02% уровня преступности, а увеличение на 1% впервые ВИЧ-инфицированных пациентов – снижение уровня преступности на 0,07%.

После проведения эконометрического моделирования и получения уравнения зависимости уровня преступности от различных факторов, можно приступать к этапу проектирования и разработки приложения. Инструментом для данного шага выступает интегрированная среда разработки MS Visual Studio, язык программирования C#, интерфейс представления Windows Forms, язык графического описания объектов - UML (Unified Modeling Language).

Приложение необходимо для прогнозирования пользователем уровня преступности путём ввода значений объясняющих переменных. Также пользователю доступны функции регулирования отдельных переменных для отслеживания динамики уровня преступности.

С помощью языка uml и диаграммы use case (вариантов использования), была спроектирована схема взаимодействия пользователя с продуктом, на основе которой были выдвинуты функциональные и нефункциональные требования к приложению. Разработка приложения включала в себя описания таких функций и методов как: обработка вводимых значений, обработка и представление данных в различных типах, переход по формам приложения, расчет значения уровня преступности по полученным данным, вывод интерпретации; на рисунке 1 представлен интерфейс работающего приложения с рассчитанным прогнозом значения уровня преступности.

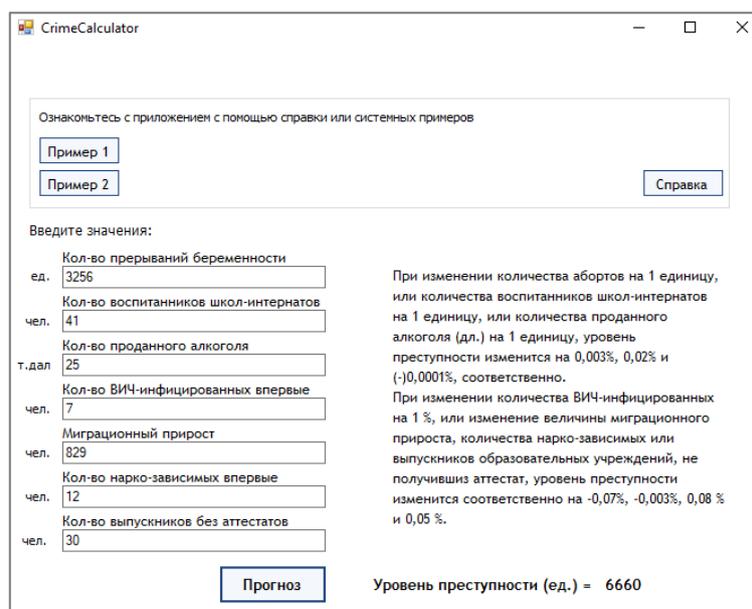


Рисунок 1. Интерфейс приложения

Проведя ряд тестов для проверки правильности выявленной модели с использованием данных 2014 - 2015 гг., была получена в среднем ошибка в 25,71% с соответствующей покрываемостью модели в 74,29%.

Таким образом, главными результатами настоящей работы являются уравнение эконометрической модели и приложение для прогнозирования уровня преступности. Полученное уравнение модели может быть в дальнейшем использоваться как основной предмет новых приложений для регулирования уровня преступности, или сравнения наиболее значимых факторов, влияющих на уровень преступности. Приложение, разработанное по полученному уравнению, может быть полезно как обучающимся в школах или университетах для реализации специальных проектов, так и государственным образованиям для решения ряда вопросов, связанных с борьбой с преступностью. Результаты исследования могут быть применены на практике студентами как технических и экономических специальностей, так и студентами исторических или юридических профилей с целью выполнения задач изучения уровня преступности в России.

Список литературы

1. Андриенко Ю.В. В поисках объяснения роста преступности в России в переходный период: криминометрический подход // Экономический журнал ВШЭ, 2001, №2, с. 194-220.
2. Носко В.П. Эконометрика. Книга вторая // Издательский дом «Дело», 2011, Москва.
3. Юдин С.В. Решение эконометрических задач и построение экономико-математических моделей с помощью пакета Gretl // Тульский государственный университет, Тула. 2008.

4. Официальные статистические показатели [Электронный ресурс] [Режим доступа: <https://fedstat.ru/>] [Проверено: 02.05.2018].
5. Kunst R. Econometric Methods for Panel Data. University of Vienna. 2010.
6. Levitt, S.D. The Impact of Legalized Abortion on Crime // The Quarterly Journal of Economics, 2001, Vol. 116, pp. 379-420.
7. Mark Duggan . More Guns, More Crime // Journal of Political Economy. 2001.
8. Paul J. Goldstein, Patricia A. Belucci. Crack and Homicide in New York City: A Case Study in the Epidemiology of Violence. From Crack in America // Demon Drugs and Social Justice, P 113-130, 1997.
9. Pop-Eleches, C. The Impact of an Abortion Ban on Socioeconomic Outcomes of Children: Evidence from Romania // Journal of Political Economy. 2006, Vol. 114.