

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ И ОБРАБОТКА ДАННЫХ

УДК 519.24

DOI: 10.18101/2304-5728-2020-2-36-51

МОДЕЛИРОВАНИЕ КРИМИНОГЕННОЙ ОБСТАНОВКИ И ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ПРЕСТУПЛЕНИЙ В РЕГИОНАХ РОССИЙСКОЙ ФЕДЕРАЦИИ

© **Аксенюшкина Елена Владимировна**
кандидат физико-математических наук, доцент,
Байкальский государственный университет
Россия, 664003, г. Иркутск, ул. Ленина, 11
aks.ev@mail.ru

© **Леонова Ольга Васильевна**
кандидат физико-математических наук, доцент,
Байкальский государственный университет
Россия, 664003, г. Иркутск, ул. Ленина, 11
olga.olgaleonova@yandex.ru

Аннотация. В работе представлена разработка комплексного эконометрического исследования криминогенной обстановки на территории Российской Федерации, в основе которого заложен регрессионный анализ. В качестве объекта исследования использовались эмпирические данные Федеральной государственной службы статистики за 2018 г. с целью выявления факторов, оказывающих значимое влияние на количество совершенных тяжких и особо тяжких преступлений в России. В процессе установления взаимосвязей между переменными был построен «веер» из шести эконометрических моделей множественной регрессии. Для выбора наилучшей модели были проведены тесты Бокса — Кокса и Зарембки, которые дали возможность выделить линейную регрессионную модель.

Полное эконометрическое исследование рассматриваемой проблемы также включало в себя анализ мультиколлинеарности факторов и исследование гетероскедастичности остатков линейной модели регрессии. Проверка неоднородности наблюдений для модели, которая в процессе исследования оказалась наилучшей из всех рассмотренных моделей, проводилась с использованием тестов Уайта, Бреуша — Пагана, Гольдфельда — Квандта, Парка и Глейзера. По всем проведенным тестам гипотеза о гомоскедастичности остатков была отклонена.

Поскольку в результате исследования предпочтение было отдано линейной регрессионной модели, то именно на основе этой модели были построены точечный и интервальный прогнозы. Установлены количественные взаимосвязи исследуемых переменных.

Ключевые слова: эконометрическое исследование; мультиколлинеарность факторов; гетероскедастичность остатков; регрессионная модель; коэффициенты эластичности; прогнозирование.

Для цитирования:

Аксеньюшкина Е. В., Леонова О. В. Моделирование криминогенной обстановки и прогнозирование количества преступлений в регионах Российской Федерации // Вестник Бурятского государственного университета. Математика, информатика. 2020. № 2. С. 36–51.

Введение

Развитие любой страны в значительной мере продиктовано воздействием неявных закономерностей большого многообразия факторов. Нахождение и исследование таких взаимосвязей для общества имеют важное значение, поскольку открывают двери для возможного управления различными показателями [2; 4; 13; 14]. Поэтому на сегодняшний день становится актуальной разработка комплексной методологической базы такого исследования на примере анализа криминогенной обстановки Российской Федерации.

В любом обществе проблема уменьшения уровня преступности всегда являлась одной из первостепенных задач. Количество совершенных преступлений является той призмой, через которую можно судить о состоянии государства, о тенденции развития общества.

Выявление факторов, непосредственно оказывающих влияние на состояние криминогенной обстановки в стране, часто поднимается в российских и международных исследованиях [7; 8; 9; 10; 11]. Несмотря на это, определение факторов, построение типа зависимости, а также проведение анализа полученных результатов еще недостаточно изучены.

В данной статье исследуется зависимость количества преступлений, совершаемых на территории Российской Федерации, от различных факторов, которые, по мнению экспертов, могут оказывать влияние на изучаемый показатель. Это дает возможность найти новые рычаги, оказывающие влияние на уровень преступности в стране.

Для проведения исследования применяются эконометрические методы. В этом случае предполагается существование связей между объясняемой и объясняющими переменными. Рассмотренный подход требует грамотного подбора всех переменных исходя из логических соображений и, что немаловажно, дает возможность отследить изменения социально-экономической обстановки в регионах Российской Федерации. Для получения результатов в работе используется регрессионный анализ [12].

Было проведено исследование статистических данных общего количества зарегистрированных тяжких и особо тяжких преступлений в Российской Федерации в 2018 г. с целью определения влияния ряда факторов на этот показатель.

1 Постановка задачи

Обозначим через Y зависимую (результатирующую) переменную — количество зарегистрированных тяжких и особо тяжких преступлений. Рассмотрим существование взаимосвязи количества преступлений со следующими независимыми факторами:

x_1 — число пациентов с психическими расстройствами, связанными с употреблением алкоголя, т. е. с синдромом алкогольной зависимости (количество зарегистрированных в год), чел;

x_2 — число пациентов с психическими расстройствами, связанными с синдромом зависимости от наркотических веществ (количество зарегистрированных в год), чел;

x_3 — средняя годовая численность постоянного населения, чел.;

x_4 — численность рабочей силы, тыс. человек;

x_5 — средний денежный доход населения, руб/чел.;

$x_6 - x_{12}$ — соотношения денежных доходов и величины прожиточного минимума, %:

x_6 — денежный доход меньше половины прожиточного минимума;

x_7 — денежный доход составляет от 0,5 до 1 прожиточного минимума;

x_8 — денежный доход составляет от 1 до 1,5 прожиточных минимумов;

x_9 — денежный доход составляет от 1,5 до 2 прожиточных минимумов;

x_{10} — денежный доход составляет от 2 до 3 прожиточных минимумов;

x_{11} — денежный доход составляет от 3 до 6 прожиточных минимумов;

x_{12} — денежный доход составляет более шести прожиточных минимумов;

x_{13} — общий прирост постоянного населения, чел.;

x_{14} — доля городского населения в общей численности населения, %;

x_{15} — число выбывших за 2018 г., чел.;

x_{16} — число прибывших за 2018 г., чел.

В качестве эмпирической базы используются открытые данные Федеральной государственной службы статистики за 2018 г.¹

Построим модели множественной регрессии — зависимости количества совершаемых преступлений от перечисленных выше факторов (при расчетах используем пакет «Анализ данных» MS Excel). Проведем анализ для выбора наилучшей модели, построим по ней точечные и интервальные прогнозы и установим количественные связи между переменными.

¹ Федеральная служба государственной статистики [Электронный ресурс]. URL: <http://www.gks.ru/> (дата обращения: 12.02.2020).

2 Построение моделей регрессии

Для построения моделей необходимо провести отбор факторов, которые будут включены в уравнение регрессии. Для отбора факторов будем использовать метод включения [3].

Для определения наиболее значимого фактора определим парные коэффициенты корреляции переменной Y с факторами $x_1 - x_{12}$. Результаты расчетов приведем в таблице 1.

Таблица 1

Парные коэффициенты корреляции

Фактор	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Коэффициент корреляции	0,712	0,854	0,960	0,967	0,391	-0,291	-0,399	-0,467
Фактор	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
Коэффициент корреляции	-0,474	-0,094	0,438	0,490	0,642	0,403	0,900	0,906

Сравнивая коэффициенты корреляции по модулю, делаем вывод, что наиболее влияющим фактором является x_4 — численность рабочей силы. На основании этого получаем следующую линейную модель для переменных Y и x_4 :

$$Y = -43,9 + 5,66x_4 + \varepsilon, \quad R^2 = 0,94, \quad (1)$$

(−0,18) (34,15)

где ε — случайная компонента, R^2 — показатель детерминации, в скобках указаны t -статистики параметров модели.

Далее в модель (1) последовательно вводим другие факторы согласно их ранжированию по коэффициенту корреляции (от большего к меньшему). Добавленные факторы тестируются на значимость с помощью критерия Стьюдента [12]. Незначимые факторы из модели исключаются и в дальнейшем исследовании не используются. В результате нескольких итераций остаются только значимые факторы: x_2 , x_4 и x_{14} , для которых множественная линейная регрессия имеет вид:

$$Y = -1800,8 + 4,7x_2 + 4,75x_4 + 25,07x_{14} + \varepsilon, \quad R^2 = 0,9495. \quad (2)$$

(−2,5) (3,59) (17,9) (2,002)

Качество модели (2) можно оценить как хорошее, поскольку все параметры модели значимые и коэффициент детерминации близок к 1.

Не будем ограничиваться только линейной моделью, рассмотрим еще и самые популярные нелинейные модели [6]: полиномиальную, полуполога-рифмическую, обратную, степенную и показательную.

Выберем наилучшую из них опытным путем, для этого проведем оценку параметров рассматриваемых моделей с помощью метода наименьших квадратов [6], предварительно проделав соответствующие преобразования. В зависимости от типа нелинейности модели будем использовать замену переменных или логарифмические преобразования.

Результаты оценок параметров моделей, *t*-статистики, коэффициенты детерминации и *F*-статистики приведены в табл. 2.

Таблица 2

Линейная и нелинейные модели в линеаризованном виде

Оцененные модели регрессии	Коэффициент детерминации, скорректированный коэффициент детерминации	Общий <i>F</i> -критерий
<p>Линейная</p> $Y = -1800,8 + 4,7 x_2 + 4,75 x_4 + 25,07 x_{14} + \varepsilon$ <p style="text-align: center;"> <small>(-2,5) (3,59) (17,9) (2,002)</small> </p>	$R^2 = 0,9495$ $\bar{R}^2 = 0,9475$	$F = 482,68$
<p>Полином второй степени</p> $Y = -765,713 + 7,9 x_2 - 0,00318 x_2^2 + 3,35 x_4 +$ <p style="text-align: center;"> <small>(-0,9) (2,25) (-0,61) (5,81)</small> </p> $+ 0,000237 x_4^2 + 9,9 x_{14} + 0,109 x_{14}^2 + \varepsilon$ <p style="text-align: center;"> <small>(2,18) (0,12) (0,18)</small> </p>	$R^2 = 0,9557$ $\bar{R}^2 = 0,9521$	$F = 266,26$
<p>Обратная</p> $\frac{1}{Y} = 0,00118 - 4,6 \cdot 10^{-7} x_2 - 6,7 \cdot 10^{-8} x_4 -$ <p style="text-align: center;"> <small>(5,7) (-1,4) (-1,04)</small> </p> $- 8,9 \cdot 10^{-6} x_{14} + \varepsilon$ <p style="text-align: center;"> <small>(-2,97)</small> </p>	$R^2 = 0,33$ $\bar{R}^2 = 0,3$	$F = 12,77$
<p>Степенная</p> $\ln Y = 0,81 + 0,19 \ln x_2 + 0,69 \ln x_4 +$ <p style="text-align: center;"> <small>(1,15) (4,88) (12,37)</small> </p> $+ 0,47 \ln x_{14} + \ln \varepsilon$ <p style="text-align: center;"> <small>(2,75)</small> </p>	$R^2 = 0,89$ $\bar{R}^2 = 0,88$	$F = 216,47$
<p>Показательная</p> $\ln Y = 6,75 + 0,001583 x_2 + 0,000359 x_4 +$ <p style="text-align: center;"> <small>(21,46) (3,27) (3,67)</small> </p> $+ 0,0114 x_{14} + \ln \varepsilon$ <p style="text-align: center;"> <small>(2,46)</small> </p>	$R^2 = 0,68$ $\bar{R}^2 = 0,67$	$F = 56,67$
<p>Полулогарифмическая</p> $Y = -35566,2 + 667 \ln x_2 + 4428,9 \ln x_4 +$ <p style="text-align: center;"> <small>(-3,92) (1,33) (6,15)</small> </p> $+ 2155,62 \ln x_{14} + \varepsilon$ <p style="text-align: center;"> <small>(0,98)</small> </p>	$R^2 = 0,61$ $\bar{R}^2 = 0,59$	$F = 40,91$

Если сравнить t -статистики параметров моделей с критической точкой $t_{кр} = 1,66$, то можно сделать вывод, что полиномиальная, обратная и полуполога-логарифмическая модели имеют незначимые параметры и в дальнейших исследованиях не участвуют.

Оставшиеся три модели являются статистически значимыми, так как их F -статистики, приведенные в таблице 2, больше, чем критическая точка $F_{кр} = 2,72$.

Таким образом, продолжим поиск модели, наиболее точно описывающей взаимосвязь между переменными среди линейной, степенной и показательной.

3 Выбор наилучшей модели

Поскольку разница между наибольшим и наименьшим значениями зависимой переменной больше 10, а также зависимые переменные в линеаризованных моделях различны, то это дает возможность использовать тест Бокса — Кокса для выбора наилучшей модели.

Тест основан на утверждении, что как исходные (наблюдаемые), так и преобразованные значения переменной являются частным случаем реализации специальной функции [5] при различных значениях λ . Суть теста заключается в поиске таких значений параметра λ , при которых остаточная сумма квадратов $SS_{ост}$ будет минимальной.

Для непреобразованного значения зависимой переменной при $\lambda = 1$ получим

$$Y^* = -1801,8 + 4,7x_2 + 4,75x_4 + 25,07x_{14} + \varepsilon; \quad SS_{ост}^{\lambda=1} = 1,43 \cdot 10^8.$$

Для логарифмированного значения зависимой переменной при $\lambda \rightarrow 0$ получим

$$Y^* = 24133,64 + 5,65x_2 + 1,28x_4 + 40,73x_{14} + \varepsilon; \quad SS_{ост}^{\lambda \rightarrow 0} = 2,5 \cdot 10^8.$$

Остаточная сумма квадратов при $\lambda = 0$ меньше, чем при $\lambda \rightarrow 0$, следовательно, связь между переменными лучше всего характеризует линейная функция в отличие от степенной и показательной.

Модификацией теста Бокса — Кокса является преобразование Зарембки [5], которое применяется только для двух форм зависимой переменной — непреобразованной и логарифмированной.

После преобразования зависимой переменной $Y^0 = \frac{Y}{\bar{Y}_{геом}}$ получены уравнения:

$$Y^0 = -0,5 + 0,0013x_2 + 0,0013x_4 + 0,007x_{14} + \varepsilon; \quad SS_{ост} = 11,2;$$
$$\ln Y^0 = -1,42 + 0,0015x_2 + 0,00035x_4 + 0,011x_{14} + \varepsilon; \quad SS_{ост} = 19,57.$$

Поскольку остаточная дисперсия у первой модели меньше, то следует отдать предпочтение линейной модели.

Проверим значимость остаточных сумм квадратов с помощью теста χ^2 . Статистика $\chi^2 = \left| \frac{81}{2} \ln \frac{11,2}{19,57} \right| = 22,5$ больше, чем критическая точка $\chi_{0,05}^2(1) = 3,84$, поэтому для всех исследуемых моделей статистически значимыми являются различия между остаточными суммами квадратов.

Таким образом, можно утверждать, что линейная модель лучше всего описывает взаимосвязь между показателями.

4 Анализ мультиколлинеарности факторов

Мультиколлинеарность факторов возникает при наличии высокой корреляционной связи между регрессорами и может привести к неустойчивости и ненадежности оценок параметров построенных моделей.

Существуют различные способы выявления и устранения мультиколлинеарности факторов [1]. Построим матрицу парных коэффициентов корреляции и проведем исследование существования корреляционной связи между факторами модели

$$r_{xx} = \begin{pmatrix} 1 & 0,82 & 0,37 \\ 0,82 & 1 & 0,35 \\ 0,37 & 0,35 & 1 \end{pmatrix}.$$

Подчеркнем, что между переменными x_2 и x_4 очень тесная связь, что подтверждает интеркорреляцию этих переменных. Между переменными x_4 и x_{14} , x_2 и x_{14} получена достаточно слабая связь. Определитель матрицы r_{xx} равен 0,27, что говорит о наличии мультиколлинеарности объясняющих переменных.

Парные коэффициенты корреляции между зависимой переменной и факторами равны $r_{Yx_2} = 0,85$; $r_{Yx_4} = 0,96$; $r_{Yx_{14}} = 0,40$. Поскольку значения межфакторной корреляции меньше этих коэффициентов, то сделаем вывод о том, что все переменные можно включить в модель.

Для измерения степени мультиколлинеарности регрессоров используем фактор вздутия дисперсии (VIF) [12]. Для этого рассчитаем коэффициенты детерминации для регрессий между переменной x_j и другими факторами:

$$\begin{aligned} \hat{x}_2 &= -81,86 + 0,16x_4 + 1,54x_{14}; & R_{x_2, x_4, x_{14}}^2 &= 0,68; \\ \hat{x}_4 &= -32,3 + 3,93x_2 + 3,84x_{14}; & R_{x_4, x_2, x_{14}}^2 &= 0,68; \\ \hat{x}_{14} &= 65,9 + 0,016x_2 + 0,0017x_4; & R_{x_{14}, x_2, x_4}^2 &= 0,14. \end{aligned}$$

Следовательно, фактор вздутия дисперсии для каждой независимой переменной будет равен

$$VIF_{x_2} = \frac{1}{1-0,68} = 3,125, \quad VIF_{x_4} = \frac{1}{1-0,68} = 3,125,$$

$$VIF_{x_{14}} = \frac{1}{1-0,14} = 1,16.$$

Рассчитанные величины свидетельствуют о наличии мультиколлинеарности переменных.

Для устранения мультиколлинеарности воспользуемся специальными методами корректировки моделей: ридж-регрессии [12] и главных компонент [1].

После применения ридж-регрессии получили уравнение:

$$y = -1733,29 + 4,71x_2 + 4,75x_4 + 24,1x_{14} + \varepsilon.$$

После применения метода главных компонент получили уравнение:

$$y = -1612,41 + 4,73x_2 + 4,75x_4 + 22,38x_{14} + \varepsilon.$$

Отметим, что изменения в моделях были относительно невелики, что указывает на невысокую степень мультиколлинеарности, которой можно пренебречь.

5 Исследование гетероскедастичности случайных остатков модели

Проверку неоднородности наблюдений в модели, которая в процессе исследования оказалась наилучшей из всех рассмотренных моделей, проведем с использованием тестов Уайта, Бреуша — Пагана, Гольдфельда — Квандта, Парка и Глейзера.

Для теста Уайта необходимо построить квадратичную функцию зависимости квадратов остатков от всех факторов и их квадратов.

Оцененная квадратичная модель имеет вид:

$$l^2 = \underset{(0,56)}{4218833} + \underset{(0,69)}{7176,3}x_2 - \underset{(-1,03)}{14,31}x_2^2 + \underset{(1,37)}{2334,29}x_4 + \underset{(0,06)}{0,02}x_4^2 - \underset{(-0,58)}{132492}x_{14} +$$

$$+ \underset{(0,5)}{861}x_{14}^2 + \delta, \quad F = 2,9,$$

где l — случайные остатки модели (1), δ — случайная компонента.

Табличное значение F -критерия равно $F_{кр} = F(0,05; 6; 74) = 2,22$. Так как $F > F_{кр}$, то по тесту Уайта гипотеза об отсутствии гетероскедастичности отклоняется.

Отметим, что все параметры этой модели незначимы, но наибольшее значение t -критерия (достаточно близкое к табличному $t_{кр} = 1,66$) имеет параметр при переменной x_4 . Таким образом, переменная x_4 может быть рассмотрена как возможная причина гетероскедастичности остатков.

Для проведения теста Бреуша — Пагана [6] необходимо оценить зависимость вида:

$$\frac{l^2}{S_1^2} = 0,83 - 0,00078x_2 + 0,001x_4 - 0,0098x_{14} + v,$$

где S_1^2 — оценка дисперсии случайных остатков, v — случайная компонента.

Наблюдаемое значение критерия $\chi^2 = \frac{SS_{ocm}}{2} = \frac{66,069}{2} = 33,034$ больше табличного $\chi_{кр}^2 = \chi^2(0,05; 3) = 7,81$, поэтому нулевая гипотеза о гомоскедастичности случайных остатков отвергается.

Исследуем влияние всех трех независимых переменных по отдельности на дисперсию случайных остатков. В итоге получили зависимости:

$$\begin{aligned} \frac{l^2}{S_1^2} &= 0,41 + 0,0032x_2 + v; & \chi^2 &= 18,09; \\ \frac{l^2}{S_1^2} &= 0,16 + 0,00087x_4 + v; & \chi^2 &= 31,955; \\ \frac{l^2}{S_1^2} &= -0,00076 - 0,014x_{14} + v; & \chi^2 &= 1,4. \end{aligned}$$

Табличное значение критерия в данном случае $\chi_{кр}^2 = \chi^2(0,05; 1) = 3,84$, неравенство $\chi^2 > \chi_{кр}^2$ справедливо только для переменных x_2 и x_4 , то есть остатки гетероскедастичны по этим переменным.

Опираясь на выявленное влияние на дисперсию остатков переменных x_2 и x_4 , проверим эту связь с помощью других тестов.

Использование теста Гольдфелда — Квандта [6] предполагает упорядочивание данных, в нашем случае по переменным x_2 и x_4 .

Общий объем наблюдений составляет 71 регион, то есть их можно разбить на три равные группы по 27 регионов. Для первой и третьей совокупности наблюдений найдем параметры множественной линейной регрессии и остаточные суммы квадратов.

Если производить упорядочивание по переменной x_2 , то получим результаты:

$$\begin{aligned} \hat{Y} &= -407,65 + 17,45x_2 + 2,76x_4 + 11,82x_{14}; & SS_{ocm}^{(1)} &= 14034165; \\ \hat{Y} &= -5202,17 + 6,44x_2 + 4,65x_4 + 59,05x_{14}; & SS_{ocm}^{(1)} &= 94056651. \end{aligned}$$

Наблюдаемое значение критерия $F = \frac{94056651}{14034165} = 6,7$. Табличное значение критерия $F_{кр} = 2,014$, так как $F > F_{кр}$, то дисперсия остатков зависит от величины значений переменной x_2 .

Если производить упорядочивание по переменной x_4 , то получим результаты:

$$\begin{aligned} \hat{Y} &= -203,62 + 8,82x_2 + 4,9x_4 + 1,8x_{14}; & SS_{ocm}^{(1)} &= 125720050; \\ \hat{Y} &= -4159,43 + 4,45x_2 + 4,79x_4 + 55,74x_{14}; & SS_{ocm}^{(1)} &= 1,5 \cdot 10^8. \end{aligned}$$

Наблюдаемое значение критерия $F = 8,38$. Табличное значение критерия $F_{кр} = 2,014$, так как $F > F_{кр}$, то дисперсия остатков зависит от величины значений переменной x_4 .

Исследование по тесту Парка [6] приводит к результатам:

$$\ln l^2 = 8,4 + 0,86 \ln x_2 + v; \quad t = 3,46,$$

$$\ln l^2 = 1,08 + 1,73 \ln x_4 + v; \quad t = 5,07.$$

Табличное значение критерия Стьюдента $t_{кр} = 1,66$; так как для обоих факторов $t > t_{кр}$, то гипотеза о гомоскедастичности отвергается.

По тесту Глейзера [6] необходимо найти параметры целой серии уравнений, задаваемых функцией

$$|\varepsilon_i| = \alpha_0 + \alpha_j x_{ji}^k + v_i,$$

где k — какое-либо число, например, $k = \pm 1; \pm 0,5$ и т. п.

Для переменной x_2 получили уравнения:

$$\text{при } k = 1, \quad |l| = 553,13 + 2,044x_2 + v, \quad t = 4,36;$$

$$\text{при } k = -1, \quad |l| = 972,39 - 1747,37 \frac{1}{x_2} + v, \quad t = -1,27;$$

$$\text{при } k = 0,5, \quad |l| = 161,95 + 65,43\sqrt{x_2} + v, \quad t = 4,59;$$

$$\text{при } k = -0,5, \quad |l| = 1208,9 - 2226,89 \frac{1}{\sqrt{x_2}} + v, \quad t = -2,43.$$

Для переменной x_4 получили уравнения:

$$\text{при } k = 1, \quad |l| = 477,72 + 0,46x_4 + v, \quad t = 5,04;$$

$$\text{при } k = -1, \quad |l| = 1175,08 - 114486,345 \frac{1}{x_4} + v, \quad t = -2,53;$$

$$\text{при } k = 0,5, \quad |l| = -183,86 + 39,4\sqrt{x_4} + v, \quad t = 5,42;$$

$$\text{при } k = -0,5, \quad |l| = 1723,93 - 18722,1 \frac{1}{\sqrt{x_4}} + v, \quad t = -3,66.$$

Табличное значение критерия Стьюдента $t_{кр} = 1,66$, таким образом, по тесту Глейзера практически при всех k гипотеза о гомоскедастичности остатков отвергается.

Подводя итоги исследования гетероскедастичности остатков изучаемых показателей, отметим, что по всем проведенным тестам гипотеза о гомоскедастичности остатков была отвергнута, то есть можно утверждать, что на дисперсию случайных остатков оказывают влияние переменные x_2 и x_4 , то есть эти переменные являются причиной гетероскедастичности остатков.

Для устранения гетероскедастичности остатков была предпринята попытка использования обобщенного метода наименьших квадратов для нахождения оценок параметров модели, но она не увенчалась успехом, поскольку некоторые параметры в новой модели оказались незначимыми.

6 Прогнозирование по линейной множественной регрессии

Результаты исследования показали, что линейная модель регрессии (2) достаточно точно описывает связь изучаемых показателей, поэтому используем ее для прогнозирования.

В таблице 3 приведены прогнозные значения переменных x_2 , x_4 и x_{14} . Используя значения этих показателей, построены точечный и интервальный прогнозы.

Таблица 3

Среднее прогнозируемое значение количества зарегистрированных тяжких и особо тяжких преступлений по федеральным округам

№	Субъект Российской Федерации	Прогнозное значение			Прогнозируемое количество преступлений (точечный прогноз)	Прогнозируемое количество преступлений (интервальный прогноз)
		x_2	x_4	x_{14}		
1	Центральный федеральный округ	95	558	65	2926	(189; 5664)
2	Московская область и г. Москва	858	5159	81	28774	(25758; 31790)
3	Северо-Западный федеральный округ	65	480	85	2916	(144; 5688)
4	Ленинградская область и г. Санкт-Петербург	363	2219	90	12704	(9921; 15488)
5	Южный федеральный округ	150	1033	53	5141	(2371; 7910)
6	Северо-Кавказский федеральный округ	93	732	56	3518	(762; 6237)

7	Приволжский федеральный округ	147	972	63	5087	(2347; 7828)
8	Уральский федеральный округ	280	1133	73	6728	(3989; 9468)
9	Сибирский федеральный округ	188	653	63	3764	(1020; 6509)
10	Иркутская область	339	1070	71	6675	(3918; 9432)
11	Дальневосточный федеральный округ	188	456	85	3380	(607; 6153)

7 Интерпретация линейной модели множественной регрессии

Построенная модель (1) дает возможность утверждать, что между количеством зарегистрированных тяжких и особо тяжких преступлений наблюдается прямая связь с числом пациентов с психическими расстройствами, связанными с употреблением алкоголя, численностью рабочей силы и долей городского населения в общей численности населения.

Коэффициенты при переменных x_2 , x_4 и x_{14} показывают абсолютную силу связи и характеризуют среднее изменение результирующего показателя при единичном изменении соответствующего фактора при условии неизменности остальных регрессоров, входящих в модель.

Таким образом, можно сделать вывод, что с изменением на одного человека числа пациентов с психическими расстройствами, связанными с синдромом зависимости от наркотических веществ, количество преступлений в среднем изменится в ту же сторону на 4,7 при неизменном уровне численности рабочей силы и доли городского населения в общей численности населения.

Изменение численности рабочей силы на 1 тыс. человек приведет к изменению количества преступлений в среднем на 4,75 при неизменных значениях числа пациентов с психическими расстройствами, связанными с синдромом зависимости от наркотических и доли городского населения в общей численности населения.

При изменении доли городского населения в общей численности населения на 1% количество преступлений в среднем изменится на 25,07 при том же уровне числа пациентов с психическими расстройствами и численности рабочей силы.

Коэффициенты регрессии нельзя сравнить между собой поскольку у них разный масштаб измерения. Для ранжирования факторов по силе их воздействия на результирующий признак можно использовать относительные показатели связи — частные коэффициенты эластичности.

Для характеристики силы связи количества преступлений и числа пациентов с психическими расстройствами, уровнем численности рабочей силы и долей городского населения в общей численности населения получены коэффициенты эластичности:

$$\bar{E}_{x_2} = 0,16; \quad \bar{E}_{x_4} = 0,85; \quad \bar{E}_{x_{14}} = 0,33.$$

То есть с изменением числа пациентов с психическими расстройствами на 1% от среднего уровня количество зарегистрированных тяжких и особо тяжких преступлений изменится в ту же сторону на 0,16% своего среднего уровня при неизменном уровне численности рабочей силы и доли городского населения в общей численности населения.

С изменением численности рабочей силы на 1% от среднего уровня количество зарегистрированных тяжких и особо тяжких преступлений изменится в ту же сторону на 0,85% своего среднего уровня при неизменном уровне числа пациентов с психическими расстройствами и доли городского населения в общей численности населения.

С изменением доли городского населения в общей численности населения на 1% от среднего уровня количество зарегистрированных тяжких и особо тяжких преступлений изменится в ту же сторону на 0,33% своего среднего уровня при неизменном уровне числа пациентов с психическими расстройствами и численности рабочей силы.

Сопоставляя полученные коэффициенты эластичности, можно сделать вывод о том, что наиболее сильное влияние на количество зарегистрированных тяжких и особо тяжких преступлений оказывает численность рабочей силы. На втором месте по силе влияния оказалась доля городского населения в общей численности населения и самое слабое влияние оказывает число пациентов с психическими расстройствами, связанными с синдромом зависимости от наркотических веществ.

Заключение

В статье использовался регрессионный анализ для моделирования и прогнозирования количества зарегистрированных тяжких и особо тяжких преступлений на основе наиболее полных и актуальных данных по восьми округам Российской Федерации. После проведенного анализа среди построенных моделей наилучшей оказалась множественная линейная регрессия. Тем не менее анализ на гетероскедастичность она не прошла, но, несмотря на это точечный прогноз, построенный по линейной модели, не сильно отличается от средних показателей преступности по каждому округу. Это говорит о том, что непостоянство дисперсии случайных остатков в данном случае не оказало значимого влияния на свойства оценок параметров. Заметим, что на точность прогнозирования влияют не только различия в количестве преступлений в различных округах, но и неоднородность других данных, например, численность рабочей силы.

Авторы считают, что построенные и верифицированные модели можно использовать в других исследовательских областях. Предложенная методика исследований открывает достаточно широкие горизонты для прогно-

зирования поведения различных социально-экономических показателей. Полученные количественные взаимосвязи между анализируемыми показателями и построенные прогнозы могут быть использованы различными службами, администрациями, правительствами областей при построении стратегии развития регионов для достижения улучшения качества жизни населения.

Литература

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы. М.: ЮНИТИ, 1998. 1000 с.
2. Балдынова Е. В., Малютина С. А. Сравнительный анализ среднемесячной реальной заработной платы работников организаций в разрезе видов экономической деятельности по Иркутской области // Известия Байкальского государственного университета. 2018. Т. 28, № 3. С. 409–418.
3. Валентинов В. А. Эконометрика. М.: Дашков и К°, 2009. 436 с.
4. Волченко Л. Ю., Мамонова Н. В., Завьялова Е. О. Моделирование влияния деятельности таможенных органов на социально-экономическое развитие и инвестиционную активность регионов // Инновационное развитие экономики. 2017. № 6(42). С. 16–26.
5. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: пер. с англ. 3-е изд. М.: Вильямс, 2007. 392 с.
6. Елисеева И. И. Эконометрика. М.: Юрайт, 2012. 449 с.
7. Кучерова С. В. Использование анализа временных рядов при исследовании уровня преступности // Фундаментальные исследования. Экономические науки. 2015. № 11. С. 1206–1209.
8. Латов Ю. В. Экономические детерминанты преступности в зарубежных странах (обзор криминометрических исследований) // Журнал институциональных исследований. 2011. Т. 3, № 1. С. 133–149.
9. Мамонова Н. В., Гаврилова Е. А. Анализ нарушения гарантий независимости адвокатов сотрудниками правоохранительных органов при защите личности в уголовном судопроизводстве // Адвокатская практика. 2019. № 2. С. 45–51.
10. Молоков В. В., Рудакова Е. Н. Исследование статистических взаимосвязей показателей преступности как фактора криминализации региона // Вестник Сибирского юридического института МВД России. 2018. № 1(30). С. 61–68.
11. Мустафина С. Ю. Влияние социально-экономических факторов на уровень преступности: статистическое исследование // Экономика инновационного развития: теория и практика. 2018. № 3 (19). С. 41–46.
12. Мхитарян В. С. Эконометрика. М.: Проспект, 2008. 380 с.
13. Рогачева О. А. Миграционные процессы в регионах Сибири и Дальнего Востока // GLOBAL AND REGIONAL RESEARCH. 2019. Т. 1, № 3. С. 256–263.
14. Трофимов С. Е. Эконометрическое моделирование динамического временного ряда цены на нефть // Известия Иркутской государственной экономической академии. 2015. Т. 25, № 6. С. 990–998.

CRIME MODELING AND FORECASTING THE NUMBER OF CRIMES
IN THE CONSTITUENT ENTITIES OF THE RUSSIAN FEDERATION

Elena V. Aksenyushkina

Cand. Sci. (Phys. and Math.), A/Prof.,
Mathematical Methods and Information Technologies Department
Baikal State University
11 Lenina St., Irkutsk 664003, Russia
aks.ev@mail.ru

Olga V. Leonova

Cand. Sci. (Phys. and Math.), A/Prof.,
Mathematical Methods and Information Technologies Department
Baikal State University
11 Lenina St., Irkutsk 664003, Russia
olga.olgaleonova@yandex.ru

Abstract. The article presents the work-out of a comprehensive econometric research of crime situation in the Russian Federation based on regression analysis. As an object of research we have used empirical data from the Federal State Statistics Service for 2018 to identify the factors that have a significant impact on the number of heavy and especially grave crimes committed in Russia. When establishing correlation between variables, we have built a "fan" of six econometric models of multiple regression. To select the best model, we have carried out Box-Cox and Zarembka tests, which made it possible to extract a linear regression model.

A complete econometric study of the problem under consideration also included the analysis of multicollinearity of factors and the study of the heteroscedasticity of the residuals of the linear regression model. Verification of the heterogeneity of observations in the model, which in the course of the study turned out to be the best of all the models considered, was carried out using the White, Breusch-Pagan, Goldfeld-Quandt, Park, and Glazer tests. For all performed tests the hypothesis of homoscedasticity of the residues was rejected.

Since, as a result of the study preference was given to a linear regression model, it was on the basis of this model that point and interval forecasts were built. Quantitative relationships of the studied variables have been established.

Keywords: econometric research; multicollinearity of factors; heteroscedasticity of residuals; regression model; elasticity coefficients; forecasting.

References

1. Aivazyan S. A., Mkhitarian V. S. *Prikladnaya statistika i osnovy* [Applied Statistics and Fundamentals]. Moscow: YuNITI Publ., 1998. 1000 p.
2. Baldynova E. V., Malyutina S. A. *Sravnitelnyi analiz srednemesyachnoi realnoi zarabotnoi platy rabotnikov organizatsii v razreze vidov ekonomicheskoi deyatelnosti po Irkutskoi oblasti* [Comparative Analysis of the Average Monthly Real Wages of Corporate Employees in Terms of Types of Economic Activity in Irkutsk Oblast]. *Izvestiya Baikalskogo gosudarstvennogo universiteta*. 2018. V. 28, no. 3. Pp. 409–418.
3. Valentinov V. A. *Ekonometrika* [Econometrics]. Moscow: Dashkov i K° Publ., 2009. 436 p.

4. Volchenko L. Yu., Mamonova N. V., Zavyalova E. O. Modelirovanie vliyaniya deyatel'nosti tamozhennykh organov na sotsialno-ekonomicheskoe razvitiye i investitsionnyuyu aktivnost' regionov [Modeling the Impact of Customs Authorities' Work on Socio-Economic Development and Investment Activity in the Regions]. *Innovatsionnoye razvitiye ekonomiki*. 2017. No. 6(42). Pp. 16–26.

5. Draper N. R., Smith H. *Applied Regression Analysis*. 3rd ed. USA: Wiley, 1998. 698 p.

6. Eliseeva I. I. *Ekonometrika* [Econometrics]. Moscow: Yurait Publ., 2012. 449 p.

7. Kucherova S. V. Ispolzovanie analiza vremennykh ryadov pri issledovanii urovnya prestupnosti [Application of Time Series Analysis in the Study of Crime Rate]. *Fundamentalnye issledovaniya. Ekonomicheskie nauki*. 2015. No. 11. Pp. 1206–1209.

8. Latov Yu. V. Ekonomicheskie determinanty prestupnosti v zarubezhnykh stranakh (obzor kriminometricheskikh issledovaniy) [Economic Determinants of Crime in Foreign Countries (review of criminometric studies)]. *Journal of Institutional Studies*. 2011. V. 3, No. 1. Pp. 133–149.

9. Mamonova N. V., Gavrilova E. A. Analiz narusheniya garantii nezavisimosti advokатов sotrudnikami pravookhranitel'nykh organov pri zashchite lichnosti v ugolovnom sudoproizvodstve [Analysis of Violations of the Guarantees of Independence of Lawyers by Law Enforcement Officials in the Protection of Personality in Criminal Proceedings]. *Advokatskaya praktika*. 2019. No. 2. Pp. 45–51.

10. Molokov V. V., Rudakova E. N. Issledovanie statisticheskikh vzaimosvyazey pokazatelei prestupnosti kak faktora kriminalizatsii regiona [Investigation of Statistical Evidence of Crime Indicators as a Factor in the Criminalization of the Region]. *Vestnik Sibirskogo yuridicheskogo instituta MVD Rossii*. 2018. No. 1(30). Pp. 61–68.

11. Mustafina S. Yu. Vliyanie sotsialno-ekonomicheskikh faktorov na uroven prestupnosti: statisticheskoe issledovanie [Impact of Socio-Economic Factors on the Crime Rate: A Statistical Study]. *Ekonomika innovatsionnogo razvitiya: teoriya i praktika*. 2018. No. 3(19). Pp. 41–46.

12. Mkhitarian V. S. *Ekonometrika* [Econometrics]. Moscow: Prospekt Publ., 2008. 380 p.

13. Rogacheva O. A. Migratsionnye protsessy v regionakh Sibiri i Dal'nego Vostoka [Migration Processes in the Regions of Siberia and the Far East]. *Global and Regional Research*. 2019. V. 1, no. 3. Pp. 256–263.

14. Trofimov S. E. Ekonomicheskoe modelirovanie dinamicheskogo vremennogo ryada tseny na nef't [Econometric Modeling of the Dynamic Time Series of Oil Prices]. *Izvestiya Irkutskoi gosudarstvennoi ekonomicheskoi akademii*. 2015. V. 25, no. 6. Pp. 990–998.