

Использование инструментария математической статистики для прогнозирования инвестиционно-инновационного потенциала Республики Башкортостан

Э.ХАЛИКОВА, Т.ЛЕЙБЕРТ, Е.БУТУСОВ

На современном этапе уровень социально-экономического развития регионов во многом определяется уровнем инвестиционного и инновационного развития, основу которого составляют интеллектуальные ресурсы, наукоемкие и информационные технологии, эффективное использование и качественное совершенствование всех факторов производства. В этой связи проблемам формирования и развития инновационного потенциала в регионах России в настоящее время уделяется значительное внимание.

На уровень инвестиционно-инновационного потенциала региона оказывают влияние не только экономические отношения, сложившиеся между субъектами экономики внутри страны, но и международные экономические отношения, природно-климатические условия, географическое расположение региона, наличие природных ресурсов на его территории, уровень технико-технологического развития отраслей народного хозяйства.

Таким образом, на обобщающий индикатор, характеризующий инвестиционно-инновационный потенциал региона, – валовой региональный продукт на душу населения региона (y), оказывают воздействие множество факторов. Это обуславливает актуальность выбора наиболее значимых из них для проведения всестороннего анализа.

Халикова Эльвира Анваровна, канд. экон. наук, старший преподаватель кафедры «Бухгалтерский учет и аудит» Уфимского государственного нефтяного технического университета (УГНТУ).
E-mail: ydacha6@yandex.ru

Лейберт Татьяна Борисовна, д-р экон. наук, заведующая кафедрой «Экономика и управление на предприятии нефтяной и газовой промышленности» УГНТУ. E-mail: lejbert@mail.ru

Бутусов Евгений Владимирович, преподаватель кафедры «Бухгалтерский учет и аудит» УГНТУ.
E-mail: butusov.e@gmail.com

Для обоснования выбора факторов, оказывающих влияние на инновационный потенциал региона, было проведено экспертное исследование, по результатам которого произведен отбор 38 показателей-факторов. В их числе – затраты на исследования и разработки, затраты на технологические инновации, число инновационно активных предприятий региона, объем инновационной продукции, инвестиции в основной капитал и др. [1].

В качестве основных факторов, оказывающих влияние на изменение валового регионального продукта, нами были отобраны следующие: x_1 – процент изношенности основного капитала; x_2 – внутренние затраты на НИОКР; x_3 – затраты на технологические инновации; x_4 – число предприятий, выполняющих НИР; x_5 – объем инвестиций в основной капитал, % к предыдущему году; x_6 – инвестиции в основной капитал на душу населения; x_7 – доля промышленной инновационной продукции в общем объеме промышленной продукции; x_8 – иностранные инвестиции, % к предыдущему году.



Рис. 1. Алгоритм проведения статистического прогнозирования экономических показателей

Как отмечает В.Г.Тихомиров, одна из основных задач создания эконометрических моделей прогнозирования экономических показателей – это получение достоверных прогнозов о дальнейшем развитии изучаемого процесса за определенный период [2]. Статистический анализ, как правило, осуществляется с применением отдельных методов математической статистики, а именно – инструментов корреляционно-регрессионного анализа.

Нами предложен алгоритм проведения статистического прогнозирования экономических показателей, который основан на последовательном применении отдельных методов математической статистики для построения наиболее достоверных и адекватных эконометрических моделей взаимосвязи индикаторов, влияющих на инвестиционно-инновационный потенциал региона (рис. 1).

Апробация предлагаемого алгоритма статистического прогнозирования экономических показателей осуществлялась на примере данных, характеризующих инвестиционно-инновационный потенциал Республики Башкортостан. Исходной информацией являлись отчетные данные Министерства экономического развития Республики Башкортостан за период с 2001 по 2012 г. (табл. 1).

Таблица 1

Исходные данные для статистического прогнозирования показателей, характеризующих инвестиционно-инновационный потенциал региона

Годы	Факторы								
	x_1 , %	x_2 , млн руб.	x_3 , млн руб.	x_4 , шт.	x_5 , %	x_6 , руб.	x_7 , %	x_8 , %	y , руб./чел.
2001	38,2	1105,05	1652,27	54	117,9	11440	2	280	82304,94
2002	38	1141,89	1635,91	56	90,48	11455	3,8	69,32	90444,99
2003	38,5	1227,84	1759,04	58	102,26	13369	1,7	195,6	113056,2
2004	40,8	1349,27	1912	60	106,01	16502	1,3	226,6	107672,6
2005	41,9	1466,6	2101,1	63	110,58	20735	3,7	270	108760,2
2006	42	1908,4	5856,7	64	113,49	26541	2,9	25,2	118217,6
2007	43,9	2783,2	5917,4	71	130,19	39551	3	521,6	124439,6
2008	43,3	3352,5	8179,5	63	109,02	50198	4,3	55,7	183168,7
2009	43,6	3480,9	11674,8	60	73,18	36453	4,5	95,1	159428,7
2010	45,4	4083	6950,1	60	99	37743	5,5	299,5	186522
2011	46,8	5413,7	13754,3	70	107,17	46337	5,6	61,9	231314
2012	50,7	7014,9	12750,3	69	108,42	57321	6,3	190,2	284068,3

Далее, в процессе исследования применение корреляционного анализа позволило определить наиболее значимые факторы, влияющие на результативный признак, а также выявить и устранить объясняющие переменные, имеющие сильную взаимосвязь друг с другом (характеризующиеся мультиколлинеарностью). В случае обнаружения сильной корреляционной связи между факторами из модели исключался тот, взаимосвязь которого с объясняемым признаком наименьшая. Для реализации корреляционного анализа была составлена матрица парных коэффициентов корреляции (табл. 2).

Таблица 2

Матрица парных коэффициентов корреляции

	y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
y	1								
x_1	0,934	1							
x_2	0,983	0,958	1						
x_3	0,885	0,863	0,912	1					
x_4	0,631	0,778	0,669	0,650	1				
x_5	-0,087	0,056	-0,061	-0,209	0,425	1			
x_6	0,898	0,918	0,909	0,888	0,760	0,070	1		
x_7	0,849	0,819	0,868	0,805	0,484	-0,263	0,789	1	
x_8	-0,196	0,021	-0,105	-0,289	0,170	0,536	-0,063	-0,225	1

Данные матрицы свидетельствуют о слабой взаимосвязи между признаком y и переменными x_4 , x_5 и x_8 , в связи с чем данные факторы не были включены в модель.

Далее была рассмотрена взаимосвязь между оставшимися объясняющими переменными: x_1 , x_2 , x_3 , x_6 , x_7 . Так как между переменными x_1 и x_2 имеется сильная зависимость, то следует исключить из модели фактор x_1 . Такой выбор объясняется тем, что данный фактор в меньшей степени зависит от результативного признака y . Аналогичным образом мы исключаем переменную x_3 , сопоставляемую с переменной x_2 . Также следует исключить фактор x_7 , так как его взаимосвязь с x_6 достаточно тесная. Таким образом, в регрессионную модель следует включать факторы x_2 и x_6 .

На втором этапе проводится регрессионный анализ. Сначала строится линейная множественная регрессия, которая в общем виде выглядит следующим образом:

$$y = a + b_1x_2 + b_2x_6 + e, \quad (1)$$

где a , b_1 , b_2 – коэффициенты регрессии; e – случайное отклонение.

В результате анализа была построена следующая модель:

$$y = 57\,072,33 + 31,09x_2 + 0,1x_6 + e. \quad (2)$$

Несмотря на высокое качество модели (коэффициент детерминации равен 0,967056565) и ее статистическую значимость в целом (F-статистика составляет 132,0977785, что больше табличного значения критерия Фишера, равного 4,256494729) коэффициент при переменной x_6 является статистически не значимым (t-статистика равна 0,182205176, в то время как табличное значение Стьюдента составляет 2,262157158).

При проведении моделирования нелинейной множественной регрессии, включающей факторы x_2 и x_6 , была получена следующая модель:

$$y = \frac{1}{-0,046 - 7,96 * 10^{-10} * x_2 + 0,046 * e^{1/x_6}} + \varepsilon \quad (3)$$

Для данной модели коэффициент детерминации составляет 0,921643216, F-статистика равна 52,92961567, и все коэффициенты регрессии являются значимыми. График нелинейной и линейной множественных регрессий, а также корреляционного поля исходных данных представлены на рисунке 2.

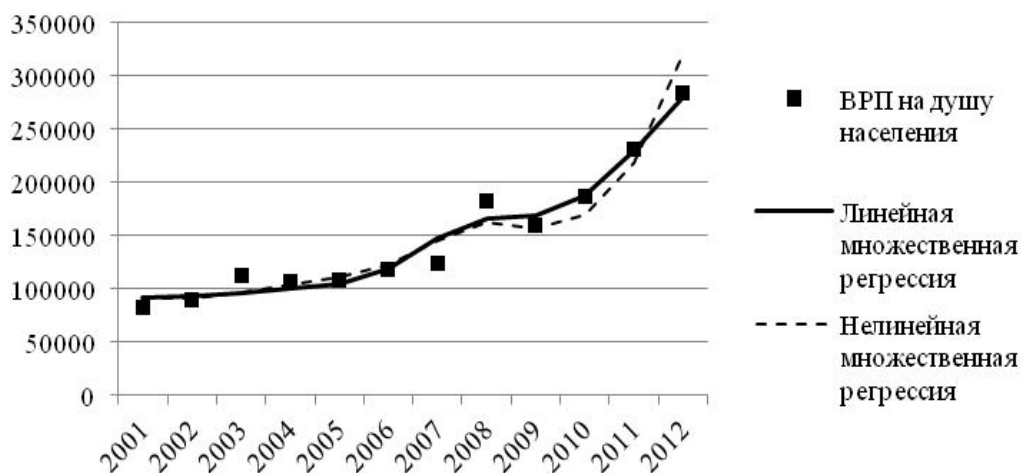


Рис. 2. Нелинейная и линейная множественная регрессия и корреляционное поле исходных данных

Несмотря на то, что для нелинейной множественной регрессии все коэффициенты являются значимыми, значение коэффициента детерминации оказалось меньше, чем для линейной. Таким образом, для дальнейшего анализа рекомендуется использовать

линейную множественную регрессионную модель. Проверка данной модели на адекватность показала положительные результаты.

Таким образом, построенная линейная множественная регрессионная модель может быть использована для определения результативного показателя на основе значений факторных переменных с достаточно высокой степенью надежности результатов. На следующем этапе для определения зависимой переменной необходимо спрогнозировать значения факторных признаков.

Для прогнозирования объясняющих факторов мы используем результаты анализа временных рядов как совокупности значений какого-либо показателя за несколько последовательных моментов или периодов. Каждый уровень временного ряда формируется из трендовой (T), циклической (S) и случайной (E) компонент. Данные компоненты представим в виде суммы $Y = T + S + E$. Задача анализа временного ряда – определить значения данных компонент.

Для нахождения количественных характеристик данных компонент мы будем использовать подход, основанный на нейронных сетях. Нейронные сети можно рассматривать как обобщение традиционных статистических подходов к анализу временных рядов. Они дают дополнительные возможности для моделирования нестационарных процессов и описания их хаотичного поведения [2].

За основу примем многослойную нейронную сеть с прямой связью. Первый слой нейронов обеспечивает определение наиболее подходящих вариантов статистической модели тренда. С учетом того, что все расчеты будут проводиться на сравнительно небольшом (часто на минимально допустимом) размере выборки, целесообразно анализировать лишь наиболее простые выражения этой компоненты.

Каждый из нейронов первого слоя будет моделировать одно из возможных представлений статистической зависимости на основе имеющейся исходной информации, а также позволит определить методом наименьших квадратов значения коэффициентов каждой из рассматриваемых функций с выдачей оценок качества построенной модели (табл. 3).

Таблица 3

Формулы представления используемых в модели нейронов [2]

Вид нейрона	Формула представления
Первый нейрон (линейная модель)	$T(t) = a + bt$
Второй нейрон (логарифмическая модель)	$T(t) = a + b \ln(t)$
Третий нейрон (показательная модель)	$T(t) = ab^t$ или $\ln(T(t)) = \ln a + t \ln(b)$
Четвертый нейрон (степенная модель)	$T(t) = at^b$ или $\ln(T(t)) = \ln a + b \ln(t)$
Пятый нейрон (гиперболическая модель)	$f(x) = a + b/t$.

Нейроны второго слоя обеспечивают определение циклической компоненты для каждого из возможных представлений тренда $f(t)$. Каждый из нейронов этого слоя будет моделировать одно из возможных представлений циклической составляющей.

С точки зрения Г.В.Тихомирова, чаще всего в качестве циклической составляющей выступает сезонная компонента, аппроксимирующая функция которой может иметь следующий вид:

$$S(t) = \frac{a}{2} + 2 \cdot \sum_{k=1}^m b_k \cdot \sin \frac{\pi kt}{6} + \sum_{k=1}^m b_k \cdot \cos \frac{\pi kt^2}{6}. \quad (4)$$

Как правило, для реальных задач $m < 3$.

В силу сделанного нами предложения об ограниченности числа наблюдений (в целях достижения стационарности случайных процессов) для того, чтобы случайные возмущения не становились статистически закономерными, будем рассматривать лишь

такие выражения циклической компоненты, которые содержат только две различные циклические переменные из выражения.

Если такое предположение неприемлемо, то есть не гарантирует достаточную степень адекватности моделирования случайного процесса, то в состав нейронов второго слоя необходимо включать алгоритмы пошаговой регрессии, и количество циклических переменных может быть увеличено. Перечень нейронов второго слоя и выражения их циклической составляющей выглядит следующим образом [2]:

$$1. S(t) = a + b_1 \cdot \sin \frac{\pi t}{6} + b_2 \cdot \sin \frac{\pi t}{3};$$

$$2. S(t) = a + b_1 \cdot \sin \frac{\pi t}{6} + b_2 \cdot \cos \frac{\pi t}{6};$$

$$3. S(t) = a + b_1 \cdot \sin \frac{\pi t}{6} + b_2 \cdot \cos \frac{\pi t}{3};$$

$$4. S(t) = a + b_1 \cdot \sin \frac{\pi t}{3} + b_2 \cdot \cos \frac{\pi t}{6};$$

$$5. S(t) = a + b_1 \cdot \sin \frac{\pi t}{3} + b_2 \cdot \cos \frac{\pi t}{3};$$

$$6. S(t) = a + b_1 \cdot \cos \frac{\pi t}{6} + b_2 \cdot \cos \frac{\pi t}{3}.$$

Таким образом, при $m = 2$ охватываются все возможные комбинации варьирования двумя циклическими переменными.

Каждый нейрон второго слоя связан со всеми нейронами первого слоя, что обеспечивает возможность полного перебора вариантов тренд-сезонных колебаний. Если обозначить $T(t)$ – представление тренда в каждом нейроне первого слоя, а $S_i(t)$ – циклическую составляющую, реализованную в i -м нейроне второго слоя, то на выходе i -го нейрона второго слоя будет тренд-сезонная компонента, имеющая следующую форму связи: $Y(t) = T(t) + S(t)$. При этом обеспечивается максимальное качество связи, характеризующееся коэффициентом детерминации R^2 .

С помощью данной нейронной сети можно оценить 30 ($5 \cdot 6$) вариантов представления тренд-сезонной составляющей для каждого временного ряда, в то время как при традиционном подходе оценке подлежали бы лишь 11 ($5+6$) вариантов представления этой составляющей.

Третий слой нейронной сети содержит только один нейрон, на вход которого подается информация о форме представления тренд-сезонной компоненты с соответствующей оценкой тесноты связи от каждого из шести нейронов второго слоя. Задача нейрона третьего слоя – сделать обоснованный выбор из шести возможных вариантов наиболее подходящего представления тренд-сезонной компоненты.

В результате проведенного анализа было определено, что наиболее качественные модели временных рядов (как для переменной x_2 , так и для переменной x_6), могут быть

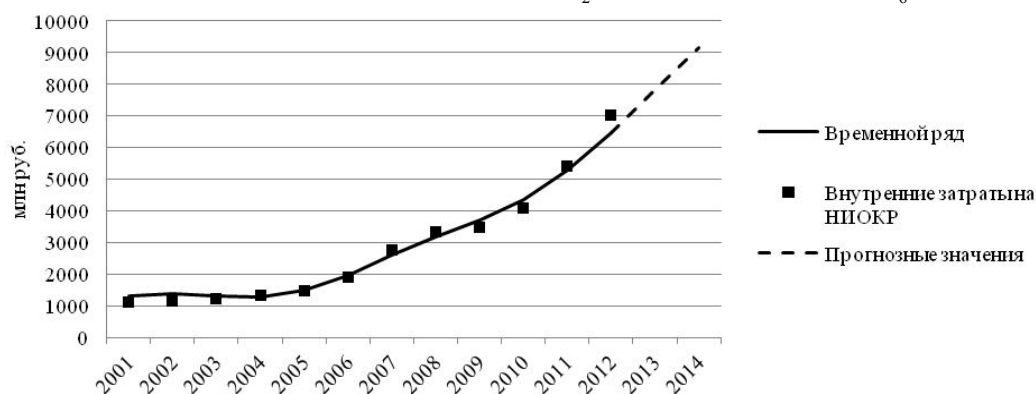


Рис. 3. График модели временного ряда и корреляционного поля для переменной x_2

построены с использованием третьего нейрона для трендовой компоненты и четвертого нейрона – для циклической. Графически модели временных рядов с исходными данными для факторов x_2 (внутренние затраты на НИОКР) и x_6 (инвестиции в основной капитал на душу населения), а также их прогнозные значения представлены на рисунках 3, 4.

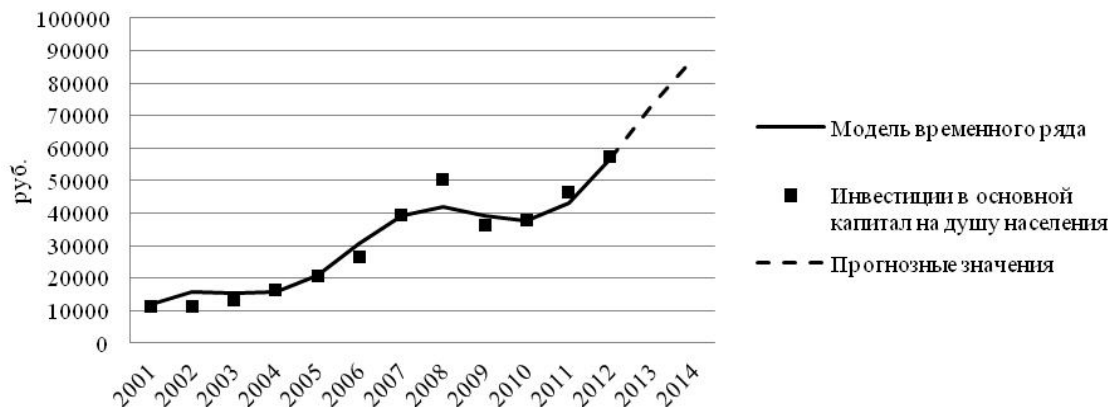


Рис. 4. График модели временного ряда и корреляционного поля для переменной x_6

На последнем этапе необходимо определить прогнозные значения показателя ВРП на душу населения, характеризующего инвестиционно-инновационный потенциал региона (рис. 5). Используя результаты прогнозирования объясняющих переменных на основе моделей временных рядов, реализованных с помощью построения нейронных сетей, можно рассчитать значения результативного признака по имеющейся нелинейной множественной регрессии.

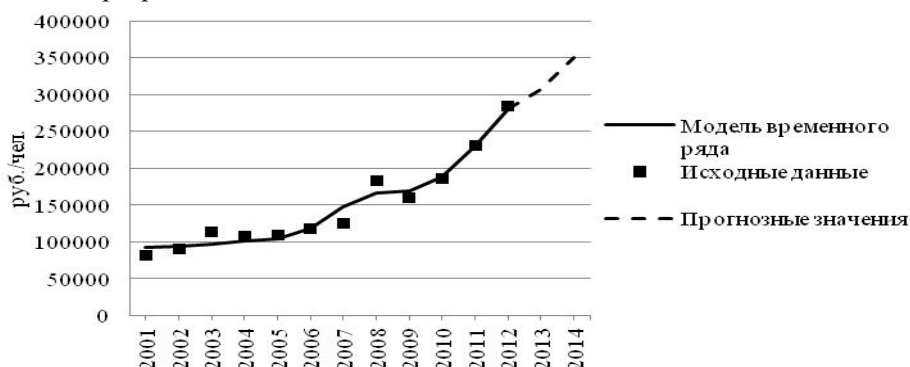


Рис. 5. График прогнозных значений на основе линейной множественной регрессии и корреляционное поля исходного ряда

Таким образом, предложенный в работе математический инструментарий прогнозирования интегрального показателя – валового регионального продукта на душу населения, характеризующего инвестиционно-инновационный потенциал региона, позволяет дать объективную оценку влияния экономических факторов на его динамику и достоверно построить его прогнозные значения на будущие периоды.

Литература

1. Шляхто И.В. Методика и результаты исследования факторов, отражающих инновационный потенциал региона // Научные ведомости Белгородского государственного университета. – 2007. – № 1. – С. 149–156.
2. Тихомиров Г.В. Прогнозирование и управление экономикой предприятия: монография. – М.: ОЛМА-ПРЕСС, 2005. – 288 с.