

МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ НАЛОГОВЫХ ПОСТУПЛЕНИЙ АДАПТИВНЫМИ МЕТОДАМИ

В статье проведен анализ динамики временного ряда налогов, выявлена зависимость между средним уровнем результативного признака и выделенными экономическими показателями и построен прогноз поступлений в бюджет Оренбургской области, который показал невозможность применения одного и того же статистического метода для планирования различных налогов.

Повышение эффективности управления бюджетным процессом требует совершенствования методов разработки и прогнозирования, прежде всего доходной части бюджета, поскольку это является определяющим фактором формирования взвешенной и обоснованной бюджетной политики. Для примера возьмем два налога: налог на прибыль организаций и налог на добычу полезных ископаемых, поскольку они играют существенную роль в доходах регионального бюджета Оренбургской области.

Визуализируем ряды поступлений налога на прибыль организаций (НП) и налога на добычу полезных ископаемых (НДПИ) [6] (рисунок 1).

Исходя из информации, представленной на рисунке 1, можно отметить, что на протяжении исследуемых периодов сохраняется положительная динамика по поступлению рассматриваемых налогов в бюджет Оренбургской области. Среди факторов, влияющих на уровень собираемости налогов, действуют факторы различной направленности: общая экономическая ситуация, период времени года и другие.

Коснемся факторов, которые непосредственно влияют на объем поступлений каждого из указанных налогов. Во-первых, это

сроки уплаты налогов, уровень инфляции, размер которой с января 2002 года по апрель 2006 года составил 52,4%, изменение долей их распределения между бюджетами, учтем, что с 1.01.2005 года доли распределения закрепились на постоянной основе ст. 56 Бюджетного кодекса РФ, что позволит делать в будущем более точные прогнозы платежей.

Во-вторых, проведем корреляционно-регрессионный анализ для выявления зависимости между средним уровнем результативного признака и выделенными экономическими показателями (НП и НДПИ от объема промышленного производства и объема добычи нефти по субъектам Приволжского федерального округа (ПФО)).

При этом предполагалось, что зависимость между результативным показателем и объясняющими переменными, отобранными для анализа, линейная, то есть наилучшая аппроксимация функции регрессии есть линейная функция, математическая модель которой имеет вид:

$$y_i = \beta_0 + \beta_1 x_{i_1} + \beta_2 x_{i_2} + \dots + \beta_k x_{i_k} + \varepsilon_i, \quad (1)$$

где n – число объектов наблюдения, $i = \overline{1, n}$, ($n=14; 10$);

y_i – значение результативного признака для i -го объекта наблюдения;

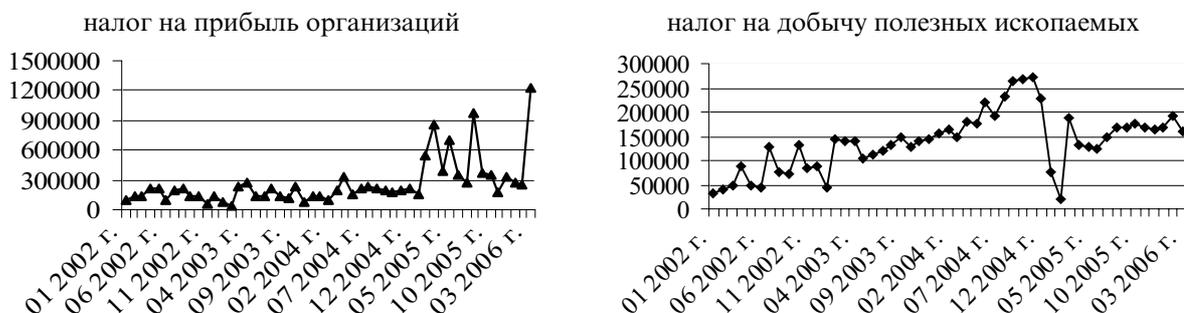


Рисунок 1. Динамика поступлений НП и НДПИ в бюджет Оренбургской области, тыс. р.

x_{ij} – значение j -й объясняющей переменной для i -го объекта наблюдения;

$\beta_0, \beta_1, \dots, \beta_k$ – подлежащие оценке параметры регрессионной модели;

ε_i – регрессионные остатки.

Предполагалось, что выполнены все условия теоремы Гаусса - Маркова [1]. Оценка коэффициентов уравнения регрессии может быть получена методом наименьших квадратов:

$$\bar{b}_{\text{МНК}} \equiv \bar{b} = (X^T X)^{-1} X^T Y, \quad (2)$$

где \bar{b} – вектор оценок коэффициентов уравнения регрессии;

X – матрица наблюдаемых значений объясняющих переменных;

Y – вектор наблюдаемых значений результативного признака.

Результативные переменные: НП и НДС. Факторные признаки – объем промышленного производства и объем добычи нефти в регионе.

При исследовании зависимостей от объема промышленного производства анализировались все субъекты ПФО (их 14); при анализе зависимости от объема добычи нефти – только те регионы округа, где есть выпуск по данной отрасли (их 10) [6, 7].

Модели регрессии строились с применением ППП «Statistica 6.0». В ходе проведения регрессионного анализа были получены следующие оценки уравнений регрессии (таблицы 1 и 2).

Регрессионные остатки данных полученных моделей распределены нормально со-

Таблица 1. Оценки уравнений регрессии зависимости от объема промышленного производства

Показатели	НП	НДС
Модель регрессии в стандартизованном масштабе	$\hat{Y} = 0,77145X$	$\hat{Y} = 0,737X$
Коэффициент детерминации R^2	59,5	54,3
F-критерий	17,64	14,26
t-критерий	4,2	3,78
Статистика Дарбина-Уотсона	2,52	1,66
Тест Колмогорова-Смирнова	0,326	0,2913

Таблица 2. Оценки уравнений регрессии зависимости от объема добычи нефти

Показатели	НП	НДС
Модель регрессии в стандартизованном масштабе	$\hat{Y} = 0,82191X$	$\hat{Y} = 0,9989X$
Коэффициент детерминации R^2	67,6	99,8
F-критерий	16,66	3508,6
t-критерий	4,08	59,23
Статистика Дарбина-Уотсона	1,242	2,06
Тест Колмогорова-Смирнова	0,218	0,2088

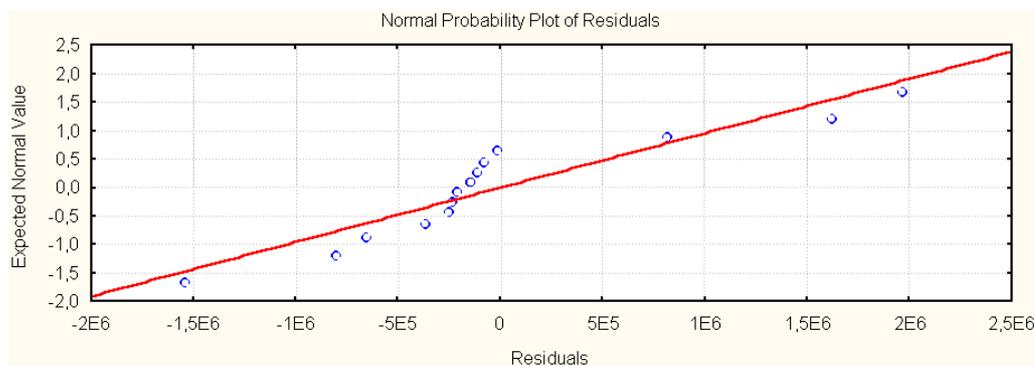


Рисунок 2. НДС – объем промышленного производства

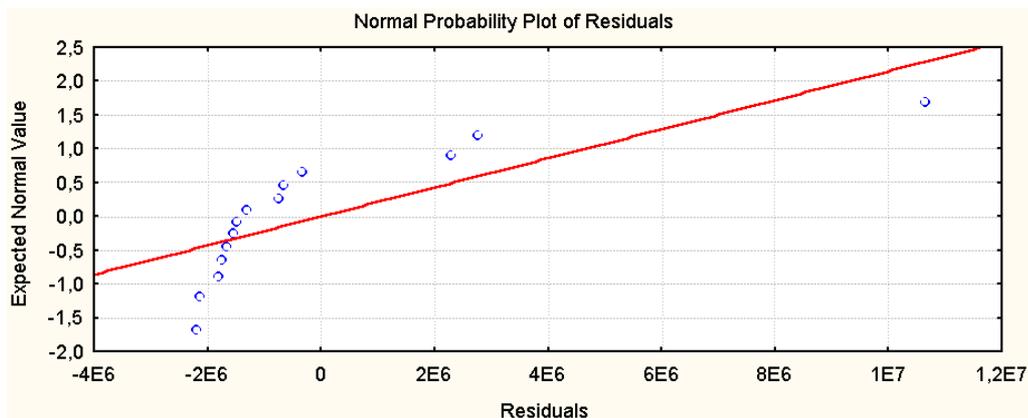


Рисунок 3. НП – объем промышленного производства

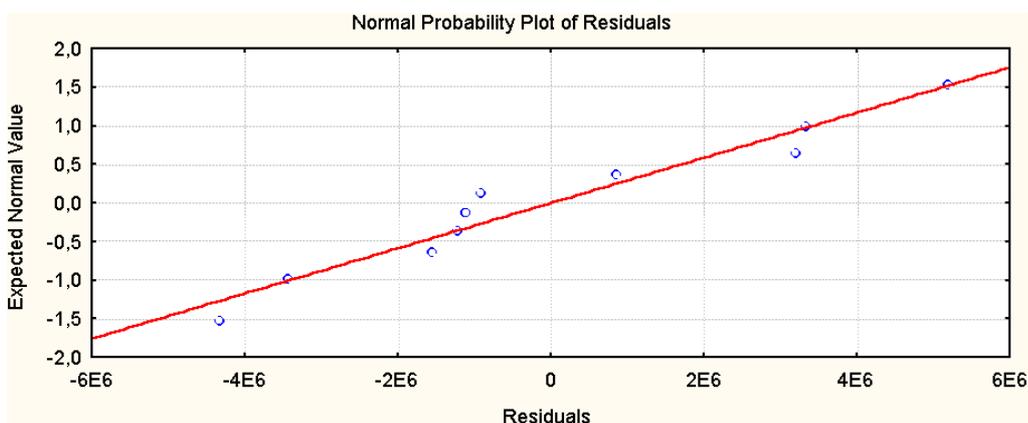


Рисунок 4. НП – объем добычи нефти

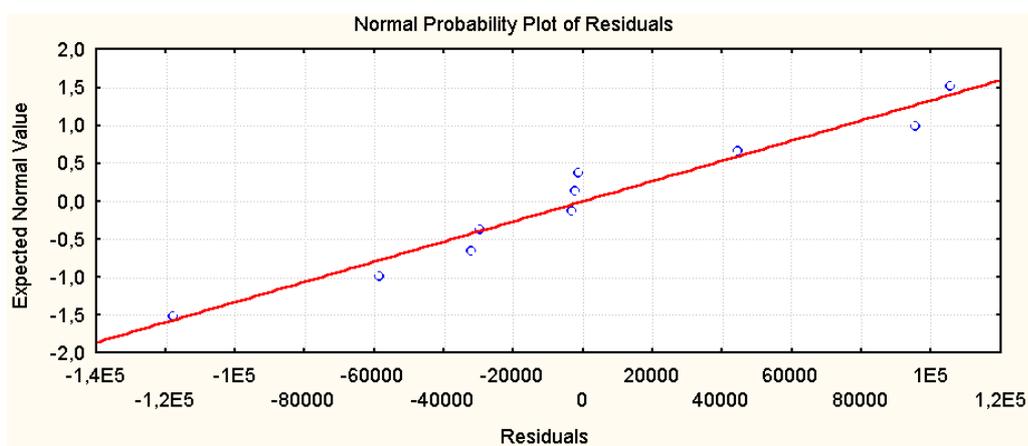


Рисунок 5. НДПИ – объем добычи нефти

гласно тесту Колмогорова - Смирнова и графикам распределения остатков на нормальной вероятностной бумаге (рисунки 2–5).

Для проверки значимости построенных моделей выдвигались гипотезы:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad (3)$$

$$H_1: \exists j \in [1, n]: \beta_j \neq 0,$$

где $k = 2$, $n = 14$.

Для проверки гипотезы H_0 использована статистика:

$$F = \frac{\hat{R}^2 / k}{(1 - \hat{R}^2) / (n - k - 1)}, \quad (4)$$

которая в случае справедливости H_0 имеет распределение Фишера - Снедекора с числом степеней свободы $v_1 = k$ и $v_2 = n - k - 1$.

По таблице распределения Фишера - Снедекора определено критическое значение статистики $F_{кр.}(\alpha = 0,05; v_1 = 2; v_2 = 11) = 3,98$ – для первых двух уравнений. $F_{кр.}(\alpha = 0,05; v_1 = 2; v_2 = 7) = 4,74$ – для вторых двух уравнений.

Поскольку $F > F_{кр.}$ во всех случаях, нулевая гипотеза отвергается, то есть построенные уравнения регрессии являются статистически значимыми.

Далее была проверена значимость отдельных коэффициентов построенных регрессионных уравнений. Выдвигались гипотезы:

H_0 : коэффициент β_j незначимо отличен от нуля ($\beta_j = 0$)

H_1 : коэффициент β_j значимо отличен от нуля ($\beta_j \neq 0$) (5)

Для проверки H_0 была использована статистика:

$$t = \frac{b_j}{S_{b_j}}, \quad (6)$$

где $j = 1, 2, \dots, k$;

b_j – оценка j -го коэффициента уравнения регрессии;

S_{b_j} – стандартная ошибка j -го коэффициента уравнения регрессии.

Данная статистика, в случае справедливости нулевой гипотезы, имеет распределение Стьюдента с $v = n - k - 1$ степенями свободы. Все коэффициенты в построенном уравнении регрессии оказались значимыми.

Если проанализировать в некоторых случаях регрессионные остатки, то можно об-

наружить, что остатки в соседних наблюдениях оказываются одного знака либо имеют разные знаки. В первом случае речь идет о наличии в остатках положительной автокорреляции, во втором – отрицательной.

Для проверки гипотезы о наличии (отсутствии) в остатках автокорреляции первого порядка использовался критерий Дарбина - Уотсона, статистика которого рассчитывается по формуле:

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^{n-1} e_i^2}. \quad (7)$$

По таблицам для заданного уровня значимости $\alpha = 0,05$ определены пороговые значения $d_n = 1,08$ и $d_u = 1,36$. Согласно расчетным значениям статистики Дарбина - Уотсона (таблицы 1 и 2), гипотеза о наличии автокорреляции в остатках отвергается.

Полученные адекватные реальным данным модели регрессии позволяют сделать следующие выводы.

Налог на прибыль организаций и налог на добычу полезных ископаемых находятся в прямой сильной зависимости от объема выпускаемой промышленной продукции (значение коэффициента R^2). Особенно тесная зависимость выявлена при моделировании регрессии НДСП от объемов добычи нефти. 99,8% вариации этого налога обусловлено именно этим показателем.

Проанализировав динамику временного ряда поступлений в бюджет области от указанных налогов за период с января 2002 года по март 2006 года и зависимость налогов от ряда факторов, моделирование рядов проведем с использованием модели АРПСС (авторегрессии – проинтегрированного скользящего среднего) Бокса - Дженкинса. Практика подтвердила гибкость данного метода.

Большинство временных рядов содержат элементы, которые последовательно зависят друг от друга. Такую зависимость можно выразить следующим уравнением:

$$X_t = \xi + \phi_1 X_{(t-1)} + \phi_2 X_{(t-2)} + \phi_3 X_{(t-3)} + \dots + \varepsilon \quad (8)$$

где ξ – константа (свободный член);

ϕ_1, ϕ_2, ϕ_3 – параметры авторегрессии.

Модели стационарных временных рядов основаны на предположении, что процесс остается в равновесии относительно постоянного среднего уровня. Однако природа устроена далеко не всегда так просто, и многие ряды, практически встречающиеся в ряде областей (таких, как индустрия, экономика, коммерция, например биржевые цены, где прогнозирование имеет особо важное значение), обнаруживают нестационарный характер, в частности не имеют фиксированного среднего. Процесс авторегрессии будет стационарным, если его параметры лежат в определенном диапазоне. Например, если имеется только один параметр, то он должен находиться в интервале $-1 < \phi < +1$. В противном случае предыдущие значения будут накапливаться и значения последующих X_t могут быть неограниченными, следовательно, ряд не будет стационарным. Если имеется несколько параметров авторегрессии, то можно определить аналогичные условия, обеспечивающие стационарность.

В отличие от процесса авторегрессии в процессе скользящего среднего каждый элемент ряда подвержен суммарному воздействию предыдущих ошибок. В общем виде это можно записать следующим образом:

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{(t-1)} - \theta_2 \varepsilon_{(t-2)} - \theta_3 \varepsilon_{(t-3)} - \dots \quad (9)$$

где μ – константа;

$\theta_1, \theta_2, \theta_3$ – параметры скользящего среднего.

Общая модель, предложенная Боксом и Дженкинсом, включает как параметры авторегрессии, так и параметры скользящего среднего. А именно имеется три типа параметров модели: параметры авторегрессии (p), порядок разности (d), параметры скользящего среднего (q). В обозначениях Бокса и Дженкинса модель записывается как АРПСС (p, d, q).

Следует отметить, что модель АРПСС является подходящей только для рядов, которые являются стационарными (среднее, дисперсия и автокорреляция примерно постоянны во времени); для нестационарных рядов следует брать разности.

Необходимо иметь достаточно большое количество наблюдений в ряде исходных данных. У нас в наличии 51 наблюдение по каждому исследуемому ряду. Проведем идентифи-

кацию, оценивание и прогнозирование исследуемых временных рядов. Приведенный выше рисунок 1 – первое свидетельство нестационарности рядов. Критерий нестационарности выражается также в отсутствии тенденции к затуханию у выборочной автокорреляционной функции ряда. Даже если бы тренд не был отчетливо виден на графике, нестационарность проявится с помощью данного критерия.

Нам необходимо сделать ряды стационарными, тогда к ним можно будет подобрать авторегрессионную модель. Применим к исследуемым рядам несколько преобразований. Для уменьшения амплитуд колебания рядов будем использовать логарифмическое преобразование. Преобразование взятия разности первого порядка позволяет избавиться от линейного тренда в ряде (смысл этого преобразования в том, что из текущего значения ряда вычитается предыдущее со сдвигом 1 и результат представляется в качестве значения нового ряда).

Отметим, что критерий стационарности носит не строгий характер, потому что в нем используются не точные автокорреляционные функции, а для их оценки, кроме того, используются не сами оценки, а графики функций, отсюда следует, что критерий допускает довольно широкое толкование и, возможно, найдется несколько приемлемых значений для порядка разности d .

На этапе идентификации определим, какое количество и каких параметров должно присутствовать в моделях. Для этого используем пакет STATISTICA. В ходе анализа автокорреляционной функции (АКФ) и частной автокорреляционной функции (ЧАКФ) преобразованных рядов наиболее приемлемыми признаны следующие модели АРПСС:

– Налог на прибыль организаций: $(0,1,1)(0,0,1)$;

– НДС: $(2,1,1)(0,1,0)$.

Оценивание параметров модели проведем приближенным методом максимального правдоподобия. Результаты оценивания представлены в таблице 3.

Ряды остатков очень похожи на белый шум. Временные ряды называются «белым шумом», если лежащая в их основе переменная имеет среднюю, равную нулю, постоянную

Таблица 3. Результаты оценивания моделей АРПСС

Параметры	Точечные оценки параметров	Асимптотическая стандартная ошибка	Значение t-критерия	Уровень значимости	Верхняя граница 95% доверительного интервала	Нижняя граница 95% доверительного интервала
Налог на прибыль организаций						
q(1)	0,772261	0,100574	7,67863	0,000	0,57004	0,974479
Qs(1)	-0,795544	0,147160	-5,40599	0,000	-1,09143	-0,499660
НДПИ						
p(1)	-0,405389	0,145462	-2,78690	0,007	-0,698189	-0,112588
p(2)	-0,305567	0,145292	-2,10313	0,041	-0,598024	-0,013110
q(1)	0,929629	0,043843	21,20337	0,000	0,841376	1,017881

дисперсию и нулевую корреляцию последовательных наблюдений, т.е. нулевую автокорреляцию. Допущения для значения остаточного члена регрессионной модели метода наименьших квадратов схожи с этим [5]. Это указывает на правильность выбранных моделей.

Подтверждением адекватности построенных моделей является также нормальность распределения остатков. Вероятность нормальности распределения остатков прослеживается на графиках нормальной плотности распределения остатков (рисунок 6). Распределение остатков симметрично и близко к нормальному. Следовательно, построенные модели среднесрочного прогнозирования для налоговых поступлений в бюджет Оренбургской области являются адекватными.

Задача прогнозирования состоит в том, чтобы по значениям наблюдений, собранных к данному моменту времени, определить значения в следующие моменты. Во всех этих ситуациях важно иметь обоснованное представление о том, что произойдет в будущем.

Статистические методы позволяют построить объективный прогноз данных. Важная особенность статистических методов прогнозирования состоит в возможности вычисления верхних и нижних границ доверительных интервалов, в которых с определенной вероятностью, например, 0,75 или 0,9, лежат значения прогнозируемых величин [2]. Знание таких границ позволяет оценить риск при принятии решения на основе прогноза.

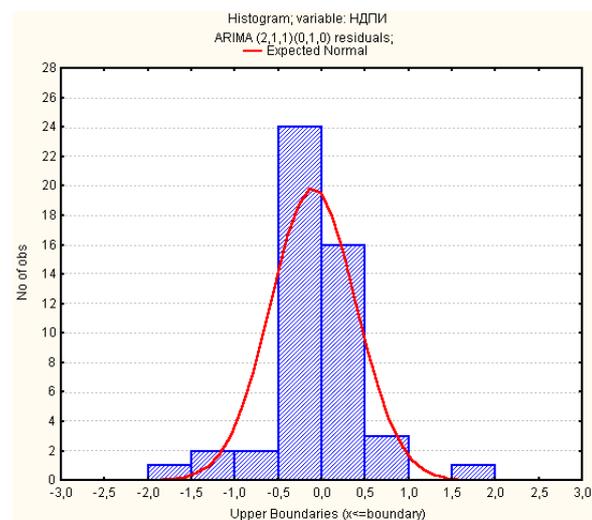
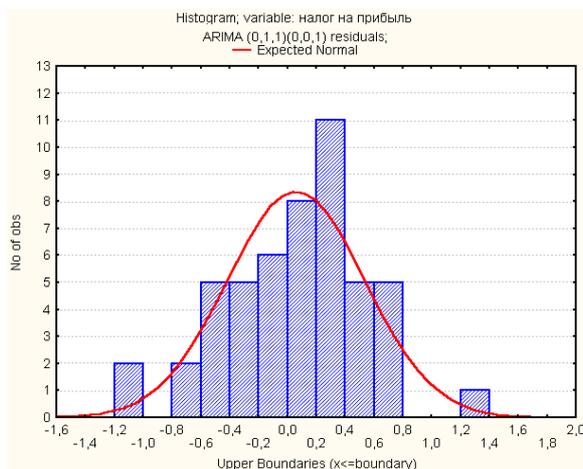


Рисунок 6. Проверка остатков моделей на нормальность и отсутствие автокорреляции

Таблица 4. Прогнозные значения по анализируемым временным рядам, в тысячах рублей

	НП			НДПИ		
	Прогноз	Нижняя граница	Верхняя граница	Прогноз	Нижняя граница	Верхняя граница
апрель 06 г.	3393340	150192	766698	170777	72385	402915
май 06 г.	604383	261975	1394328	180753	64477	506720
июнь 06 г.	391110	166113	920862	178393	57086	557477
июль 06 г.	296280	123358	711599	181637	48474	680611
август 06 г.	683173	278966	1673053	186561	42212	824525
сентябрь 06 г.	375189	150316	936473	189133	36897	969498
октябрь 06 г.	484688	190599	1232552	192241	32006	1154707
ноябрь 06 г.	242738	93725	628666	195979	27799	1381611
декабрь 06 г.	334004	126672	880689	199390	24140	1646925
январь 07 г.	414310	154385	1111846	202842	20906	1968064
февраль 07 г.	211246	77367	576795	206487	18077	2358690
март 07 г.	434365	156399	1206352	210147	15603	2830297

В ходе проведенного исследования мы получили удовлетворительные результаты оценивания параметров; анализ остатков позволил определить, что построенные модели достаточно адекватно описывают наблюдаемые временные ряды. Следовательно, с определенной степенью вероятности можно доверять прогнозам, построенным с помощью подобранных моделей. Уровень (коэффициент) доверия, измеряющий надежность прогнозируемых значений, зададим равным 0,9. На основе полученных моделей исчислены прогнозные данные индексов на среднесрочную перспективу (за период с апреля 2006 г. по март 2007 г.). Чем меньше ширина полученных в моделях доверительных интервалов, тем меньше риск при принятии решения на основе прогноза.

По результатам таблицы 4 необходимо отметить, что не для всех налоговых поступ-

лений могут быть применимы одни и те же статистические методы планирования. Однако модель АРПСС не следует совсем отвергать, поскольку рассчитываются верхняя и нижняя границы доверительного интервала, куда возможно попадание фактического варианта, который может иметь большее или меньшее значение по сравнению с рассчитанным показателем прогноза.

Наличие достоверных данных по планируемым показателям налоговых доходов дает информацию региональным органам исполнительной и законодательной власти для выработки на очередной прогнозируемый период конкретных мер региональной экономической политики, неотъемлемой частью которой являются вопросы инвестиционной политики и поддержки отдельных отраслей промышленности и экономики в целом.

Список использованной литературы:

1. Айвазян С.А. Прикладная статистика и основы эконометрики: учебник для вузов / С.А. Айвазян, В.С. Мхитарян. – М.: ЮНИТИ-ДАНА, 2001. – 1022 с.
2. Боровиков В.П., Ивченко Г.И. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере: учебное пособие. – М.: Финансы и статистика, 1999. – 384 с.
3. Домбровский В.В. Эконометрика: учебник. – М.: Новый учебник, – 2004. – 342 с.
4. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: учебник. – М.: Финансы и статистика, 2000. – 352 с.
5. Курс социально-экономической статистики: учебник для вузов / Под ред. Назарова М.Г. – М.: Финстатинформ, 2002. – 976 с.
6. Официальный сайт Федерального казначейства России // www.roskazna.ru
7. Российский статистический ежегодник, 2005: Статистический сборник / Росстат, М.: 2005. – 725 с.
8. Уотшем Т. Дж., Паррамоу К. Количественные методы в финансах: учебное пособие для вузов / Пер. с англ. Под ред. Ефимовой М.Р. – М.: Финансы, Юнити, 1999. – 527 с.