

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2020, №1, Том 12 / 2020, No 1, Vol 12 <https://esj.today/issue-1-2020.html>

URL статьи: <https://esj.today/PDF/05ECVN120.pdf>

Ссылка для цитирования этой статьи:

Шумилина В.Е., Цвиль М.М. Построение модели регрессии по временным рядам с целью прогнозирования индекса производительности труда в Российской Федерации // Вестник Евразийской науки, 2020 №1, <https://esj.today/PDF/05ECVN120.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

For citation:

Shumilina V.E., Tsvil M.M. (2020). Building a time series regression model with the aim of predicting the labor productivity index in the Russian Federation. *The Eurasian Scientific Journal*, [online] 1(12). Available at: <https://esj.today/PDF/05ECVN120.pdf> (in Russian)

УДК 330.3; 311.2

ГРНТИ 06.35.33

Шумилина Вера Евгеньевна

ФГБОУ ВО «Донской государственный технический университет», Ростов-на-Дону, Россия
Доцент

Кандидат экономических наук

E-mail: Shumilina.vera@list.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=510410

Цвиль Мария Михайловна

ГКОУ ВО «Российская таможенная академия»

Филиал в г. Ростов-на-Дону, Ростов-на-Дону, Россия

Доцент

Кандидат физико-математических наук, доцент

E-mail: tsvilmm@mail.ru

РИНЦ: https://www.elibrary.ru/author_profile.asp?id=768928

Построение модели регрессии по временным рядам с целью прогнозирования индекса производительности труда в Российской Федерации

Аннотация. В рамках Стратегии экономической безопасности РФ до 2030 года, принятой Указом Президента РФ в 2017 году, одним из отслеживаемых показателей является индекс производительности труда, который представляет собой важный индикатор экономической эффективности производства, реализации трудового и человеческого потенциала, развития техники и технологии.

Целью данного исследования является прогнозирование индекса производительности труда по Российской Федерации на последующие периоды. В процессе исследования продемонстрирована специфика взаимосвязи по рядам динамики при составлении модели регрессии. В качестве временных рядов рассматриваются динамика инвестиций в основной капитал (x_t) и изменение индекса производительности труда (y_t) в РФ за период 2008–2017 гг. В ходе исследования ранее полученного уравнения регрессии $y(x)$ выявлено наличие автокорреляции в остатках ($y - \hat{y}_x$). Было установлено, что данные ряды динамики характеризуются наличием тренда. Рассматриваемые в ходе исследования модели регрессии рядов динамики x_t , y_t имеют полиномиальные тренды четвертой степени, улучшенные с помощью фиктивных переменных. При построении модели регрессии по данным временным рядам возникла необходимость учета тенденции. Был выбран метод учета тенденций – метод

отклонений от тренда. В этом случае модель регрессии строится по остаточным величинам. С этой целью по каждому из рядов находятся остаточные величины d_y и d_x . В ходе исследования получена статистическая модель зависимости индекса производительности труда от динамики инвестиций в основной капитал в целом по Российской Федерации, позволяющая сделать прогноз индекса производительности труда на последующие периоды.

Ключевые слова: производительность труда; прогнозирование; моделирование; инвестиции в основной капитал; регрессионный анализ; полином четвертой степени; коэффициент детерминации; автокорреляция; тренд

Правительство Российской Федерации (РФ) ежегодно говорит о повышении производительности труда в стране. Данный показатель является одним из отслеживаемых в рамках Стратегии экономической безопасности Российской Федерации до 2030 года, поэтому анализ и прогнозирование индекса производительности труда в стране является актуальной темой исследования [1–3].

В нашем предыдущем исследовании [3] в данном направлении, перебирая многочисленные варианты, мы пришли к выводу, что среди рассмотренных переменных (динамика инвестиций в основной капитал; степень износа основных фондов; реальная начисленная заработная плата; число прибывших в РФ мигрантов; инновационная активность организаций; коэффициент изобретательской активности) основным фактором, влияющим на индекс производительности труда в РФ, является динамика инвестиций в основной капитал.

Целью данного исследования является статистический анализ и моделирование индекса производительности труда под влиянием динамики инвестиций в основной капитал для дальнейшего его прогнозирования.

Для проведения нашего исследования мы использовали официальные данные Федеральной службы государственной статистики за период 2008–2017 гг., которые были доступны по состоянию на декабрь 2019 г. На первом этапе исследования в качестве эндогенной переменной выступает y – индекс производительности труда в РФ (в процентах к прошлому году) (см. таблицу 1). В качестве объясняющей (экзогенной) переменной использовалась x – динамика инвестиций в основной капитал (в процентах к прошлому году) по РФ (см. таблицу 1):

Таблица 1

Данные для статистического моделирования индекса производительности труда в РФ¹

Годы	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
x	109,5	86,5	106,3	110,8	106,8	100,8	98,5	89,9	99,8	104,8
y	104,8	95,9	103,2	103,8	103,3	102,2	100,7	98,1	99,7	101,5

С помощью инструментария «Анализа данных» программного пакета Excel в предыдущем нашем исследовании [3] было получено уравнение линейной парной регрессии

$$y = 67,69 + 0,332x_1 \quad (1)$$

Коэффициент детерминации составил 0,93. Казалось, что мы достигли поставленной цели, получив уравнение регрессии (1), пригодное для прогноза. Однако, по результатам изучения остатков $e = y - \hat{y}$ было обнаружено наличие автокорреляции.

¹ Составлено авторами на основе официальных статистических данных Федеральной службы государственной статистики www.gks.ru.

Наличие автокорреляции между соседними членами, свидетельствующее о том, что на последующее наблюдение оказывает наибольшее воздействие значение предыдущего, позволяет выявить тест Дарбина-Уотсона [4; 6]. Он основан на положении, что, если корреляция ошибок регрессии не равна нулю, то она имеет место и в остатках регрессии e_t , которые получаются как результат применения метода наименьших квадратов. В тесте Дарбина-Уотсона для оценки корреляции применяется статистика вида

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (2)$$

Далее для d – статистики определяются два числа d_1 и d_2 по таблице критических точек Дарбина-Уотсона, которые имеют зависимость лишь от числа наблюдений, числа регрессоров и уровня значимости $\alpha = 0,05$: $d_1 = 0,979$, $d_2 = 1,32$.

Изобразим результат Дарбина-Уотсона графически (рисунок 1) [4; 6]:

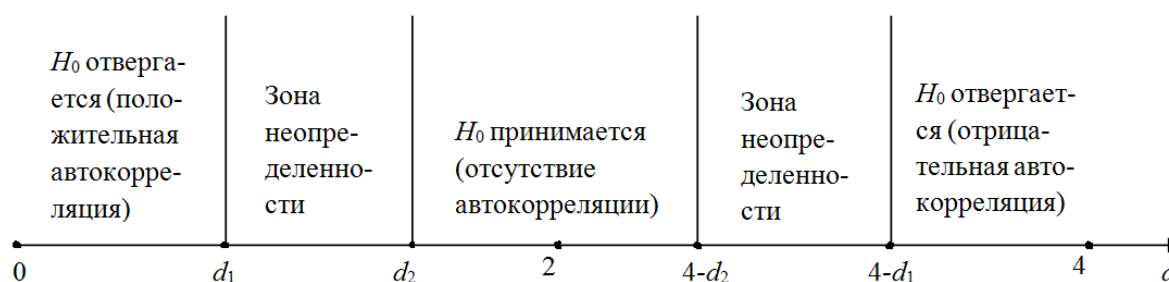


Рисунок 1. Интервалы для критерия Дарбина-Уотсона

Фактическое значение для нашей модели, согласно формуле (2), $d = 3,11$ и оно попадает в промежуток от $4 - d_1$ до 4. Следовательно, автокорреляция в остатках присутствует.

Делаем вывод, что полученное ранее уравнение регрессии не учитывает влияние тенденций, а потому содержит систематическую погрешность.

Высокое значение коэффициента детерминации показывает, что ряды $x(t)$, $y(t)$ имеют тенденцию повышения уровней. При этом оценка параметра регрессии окажется неэффективной.

В результате требуется предварительная специальная обработка рядов динамики при построении по ним регрессионных моделей. В случае наличия тренда в рядах динамики построение модели регрессии должно сопровождаться его исключением.

На основе данных таблицы 2 изучим более подробно временные ряды динамики $x(t)$, $y(t)$ и построим по ним модель регрессии с учетом того, что каждый из этих рядов обладает тенденцией.

Таблица 2

Ряды динамики $x(t)$, $y(t)$ ¹

t	1	2	3	4	5	6	7	8	9	10
x	109,5	86,5	106,3	110,8	106,8	100,8	98,5	89,9	99,8	104,8
y	104,8	95,9	103,2	103,8	103,3	102,2	100,7	98,1	99,7	101,5

По данным таблицы 2 построим график зависимости $x(t)$ (рисунок 2).

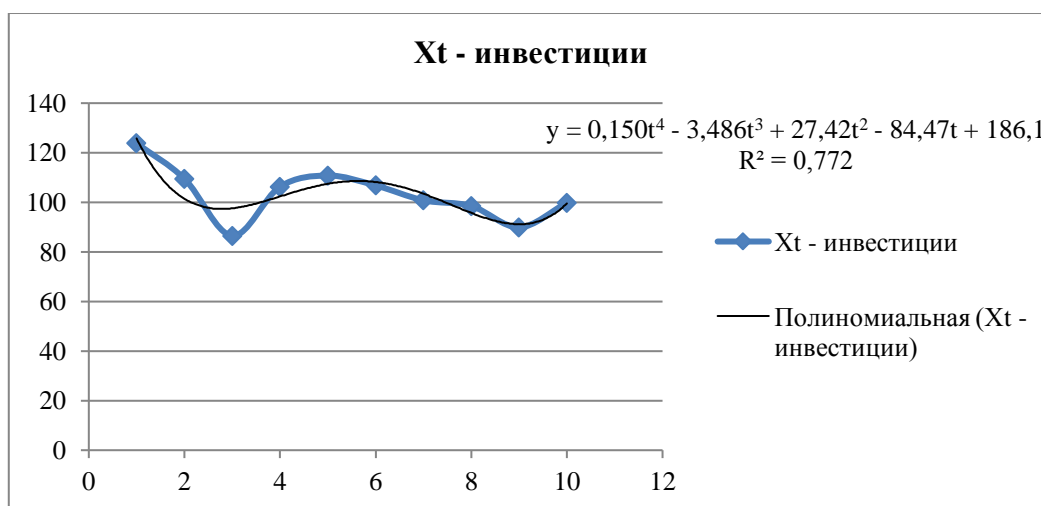


Рисунок 2. График зависимости $x(t)$ по данным таблицы 2 (получено авторами)

Из графика зависимости $x(t)$ следует, что уравнение тренда будем искать в виде полинома четвертой степени. Проведя соответствующую замену переменных, применяем программу «Регрессия» из Пакета анализа приложения MS Excel и, используя фиктивную переменную $z_3 = (0,0,1,0,0,0,0,0,0,0)$ и $z_8 = (0,0,0,0,0,0,0,1,0,0)$, получим улучшенную модель вида

$$\hat{x}_t = 170,75 - 0,0925t^3 + 0,009t^4 - 23,65z_3 - 8,81z_8 \quad (3)$$

Суть фиктивных переменных состоит в том, что они имеют только два значения – единица и ноль, поэтому их определяют как переменные бинарного типа. Мы используем фиктивные переменные в данном уравнении для отражения влияния значительных колебаний (отклонений) на резульативный показатель.

Результаты регрессионной статистики и дисперсионного анализа для модели (3) представлены на рисунке 3.

1	ВЫВОД ИТОГОВ					
2						
3	<i>Регрессионная статистика</i>					
4	Множественн	0,973626706				
5	R-квадрат	0,947948962				
6	Нормированн	0,906308132				
7	Стандартная с	2,466269744				
8	Наблюдения	10				
9						
10	<i>Дисперсионный анализ</i>					
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12	Регрессия	4	553,8685678	138,4671419	22,76489115	0,002082989
13	Остаток	5	30,41243224	6,082486448		
14	Итого	9	584,281			
15						
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95% v:</i>
17	Y-пересечени	110,7487413	1,652334707	67,02560976	1,3998E-08	106,5012797
18	t^3	-0,092545414	0,023946436	-3,864684323	0,011823736	-0,154101686
19	t^4	0,008629958	0,002375842	3,632377994	0,015023036	0,00252266
20	z2	-23,64645727	2,907234018	-8,133661455	0,000456027	-31,11974023
21	z8	-8,813795955	2,967562954	-2,970045149	0,031155469	-16,44215938

Рисунок 3. Результаты регрессионной статистики по модели (3) (получено авторами)

Автокорреляция в остатках отсутствует.

Сравним графическое представление фактических значений инвестиций в основной капитал $x(t)$ с предсказанными по модели (3) (рисунок 4).

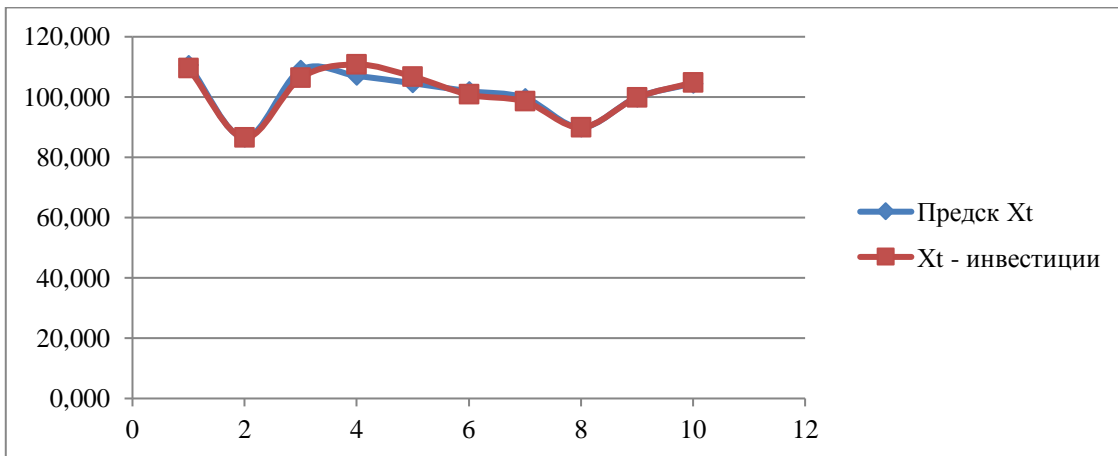


Рисунок 4. Графическое представление фактических значений инвестиций в основной капитал $x(t)$ и предсказанных (получено авторами)

Далее по данным таблицы 2 построим график зависимости y от t (рисунок 5).

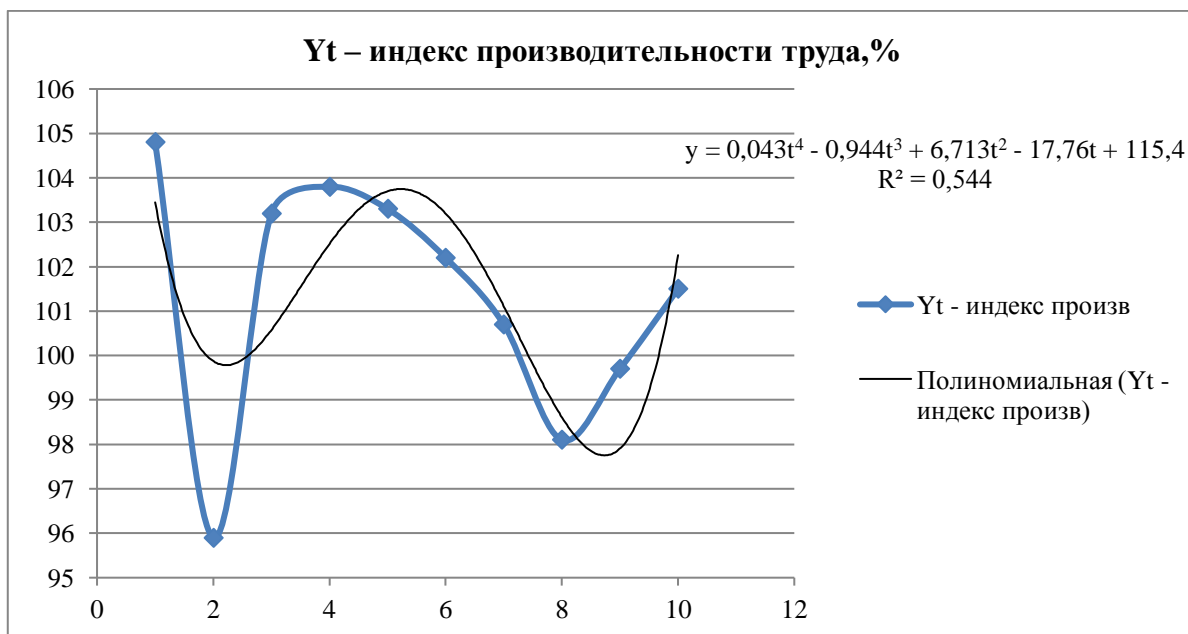


Рисунок 5. График зависимости по данным таблицы 2 (получено авторами)

Из графика зависимости $y(t)$ следует, что уравнение тренда будем искать в виде полинома четвертой степени. Проведя соответствующую замену переменных, применяем программу «Регрессия» из Пакета анализа приложения MS Excel в результате многочисленных попыток использования различных фиктивных переменных, получим улучшенную модель вида

$$\hat{y}_t = 108,35 - 5,3t + 2t^2 - 0,3t^3 + 0,015t^4 - 7,67z_2 - 1,59z_8, \quad (4)$$

где $z_2 = (0,1,0,0,0,0,0,0,0,0)$; $z_8 = (0,0,0,0,0,0,0,1,0,0)$.

Результаты регрессионной статистики и дисперсионного анализа для модели (4) представлены на рисунке 6. Автокорреляция в остатках отсутствует.

Сравним графическое представление фактических значений индекса производительности труда $y(t)$ с предсказанными по модели (4) (рисунок 7).

1	ВЫВОД ИТОГОВ					
2						
3	Регрессионная статистика					
4	Множественный коэффициент	0,998111583				
5	R-квадрат	0,996226733				
6	Нормированный коэффициент	0,988680199				
7	Стандартная ошибка	0,295181779				
8	Наблюдения	10				
9						
10	Дисперсионный анализ					
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12	Регрессия	6	69,01460315	11,50243386	132,0111617	0,001009453
13	Остаток	3	0,261396848	0,087132283		
14	Итого	9	69,276			
15						
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
17	Y-пересечение	108,3475117	1,050623165	103,1269016	2,01005E-06	105,0039599
18	t	-5,300750477	1,292669918	-4,100621823	0,026239936	-9,414603082
19	t^2	2,008994929	0,462226651	4,346341619	0,022488513	0,537983433
20	t^3	-0,307423513	0,062727479	-4,900938438	0,016258838	-0,507050347
21	t^4	0,015277234	0,002859968	5,341750049	0,012828378	0,00617554
22	z2	-7,667038116	0,407647578	-18,80800606	0,000328126	-8,964354644
23	z8	-1,591895657	0,391787911	-4,063156649	0,026881136	-2,838739647

Рисунок 6. Результаты регрессионной статистики по модели (4) (получено авторами)

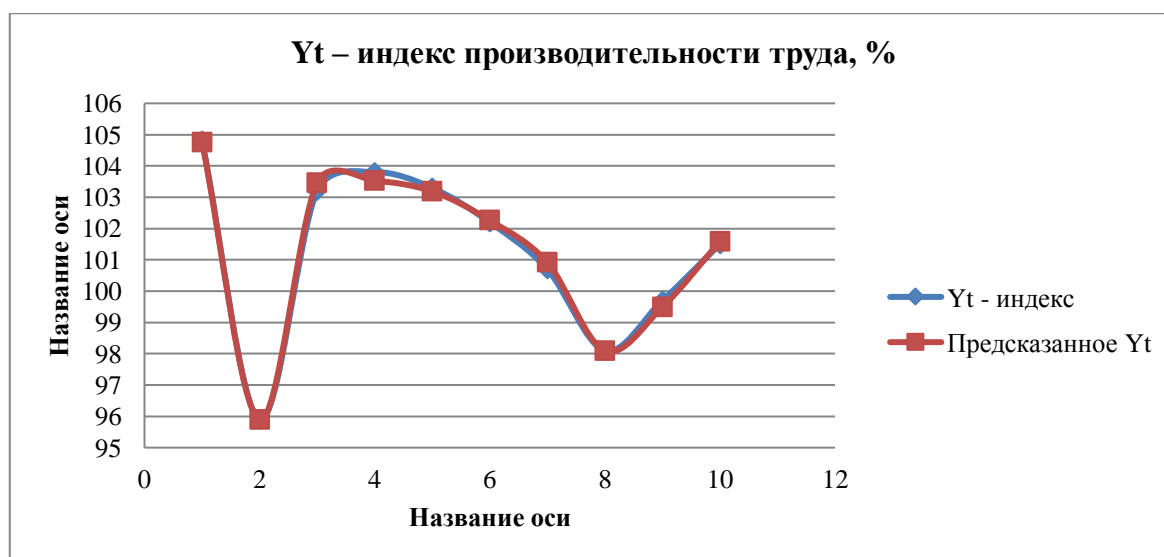


Рисунок 7. Графическое представление фактических значений индекса производительности труда $y(t)$ и предсказанных (получено авторами)

Определение присутствия в каждом из рядов $x(t)$, $y(t)$ четко прослеживающейся тенденции указывает на то, что результаты регрессионно-корреляционного анализа могут быть завышены. В данном случае целесообразно использовать метод исключения тенденций из

данных временных рядов для построения модели регрессии по временным рядам $x(t)$, $y(t)$. С этой целью применяем метод отклонений от тренда. Найдем $dy = y_t - \hat{y}_t$ и $dx = x_t - \hat{x}_t$ (таблица 3).

Таблица 3

Ряды отклонений dy , dx

dx	-2,649	-1,226	-1,165	-1,143	-0,104	0,000	0,000	0,297	2,226	3,765
dy	-0,263	-0,217	0,036	-0,063	0,208	0,000	0,000	-0,088	0,111	0,276

Получено авторами

Применяя к рядам отклонений dy , dx метод наименьших квадратов, получим уравнение линейной регрессии (4):

$$dy = -0,000000000000005 + 0,07 \cdot dx. \quad (5)$$

При этом коэффициент детерминации $R^2 = 0,6$. Полученное уравнение регрессии (5) свидетельствует о том, что при исключении из первоначальных уровней временных рядов тренда присутствует связь между остаточными величинами. Однако, качество полученного уравнения не позволяет использовать его для достоверного прогноза. По этой причине лучшим модель, снова используя фиктивные переменные.

В результате многократных попыток использования фиктивных переменных получено качественное уравнение линейной регрессии (6).

$$dy = -0,05 + 0,08dx + 0,19z_3 + 0,09z_4 + 0,27z_5, \quad (6)$$

где $z_3 = (0,0,1,0,0,0,0,0,0)$, $z_4 = (0,0,0,1,0,0,0,0,0)$, $z_5 = (0,0,0,0,1,0,0,0,0)$, $R^2 = 0,95$. Приведем результаты регрессионной статистики (рисунок 8).

Регрессионная статистика						
4	Множеств	0,973245143				
5	R-квадрат	0,947206109				
6	Нормиро	0,904970996				
7	Стандарт	0,052536001				
8	Наблюда	10				
9						
Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
12	Регрессия	4	0,247596691	0,061899173	22,4269818	0,002156916
13	Остаток	5	0,013800157	0,002760031		
14	Итого	9	0,261396848			
15						
Кoeffициенты Стандартная ошибка t-статистика P-Значение Нижние 95% df						
17	Y-пересеч	-0,055305452	0,020161206	-2,743161814	0,040639028	-0,107131483
18	z3	0,1906429	0,058206951	3,275260048	0,022069818	0,04101717
19	z5	0,272625414	0,056347094	4,838322554	0,004721886	0,127780599
20	dx	0,084946035	0,010128196	8,387084463	0,000394707	0,058910679
21	z4	0,089786204	0,058150218	1,544038989	0,183231884	-0,05969369

Рисунок 8. Результаты регрессионной статистики по модели (6) (получено авторами)

Поэтому модель (6) пригодна для использования ее в процессе прогнозирования, что подтверждается отсутствием в ней автокорреляции остатков. Используя формулу (2) – статистику Дарбина-Уотсона – получаем $d = 1,98$. Используя таблицу критических точек Дарбина-Уотсона, определяем значения d_1 и d_2 , которые зависят лишь от числа наблюдений, числа регрессоров и уровня значимости $\alpha = 0,05$: $d_1 = 0,376$, $d_2 = 2,414$. Наличие автокорреляции не подтверждается.

Для прогноза на предстоящий период воспользуемся уравнением вида:

$$y_p = \hat{y}_{t=p} - 0,05 + 0,08 (x_p - \hat{x}_{t=p}), \quad (7)$$

где $\hat{y}_{t=p} = 118,365$, $\hat{x}_{t=p} = 179,4015$.

Тогда y_p составит

$$y_p = 118,365 - 0,05 + 0,08(11 - 179,4015) = 104,84.$$

Полученная модель (7) характеризует изменения индекса производительности труда в зависимости от динамики инвестиций в основной капитал. Значимость данного исследования состоит в том, что эту модель можно использовать для составления прогнозов индекса производительности труда в РФ на предстоящие периоды.

ЛИТЕРАТУРА

1. Лядова Е.В. Анализ динамики производительности труда в России: макроэкономический аспект // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2017. №1 (45). [Электронный ресурс]. Режим доступа: <https://cyberleninka.ru/article/n/analiz-dinamiki-proizvoditelnosti-truda-v-rossii-makroekonomicheskiy-aspekt>, свободный. – Загл. с экрана. – Яз. рус.
2. Смирнова Е.А., Хохлова Н.В. Прогнозирование производительности труда в сельскохозяйственных предприятиях Ульяновской области // Молодой ученый. – 2014. – №3. – С. 545–547. – Режим доступа: <https://moluch.ru/archive/62/9496/> свободный. – Загл. с экрана. – Яз. рус.
3. Цвиль М.М., Шумилина В.Е., Статистическое моделирование и прогнозирование индекса производительности труда в Российской Федерации // Вестник Евразийской науки, 2019 №1, Том 11 [Электронный ресурс]. Режим доступа: <https://esj.today/PDF/63ECVN119.pdf>, свободный. – Загл. с экрана. – Яз. рус.
4. Цвиль М.М. Анализ временных рядов и прогнозирование: учеб. пособие. Ростов н/Д: РГА, Ростовский филиал. – 2016. – с. 135.
5. Цвиль М.М., Шумилина В.Е. Применение моделей анализа панельных данных для оценки объема инновационных товаров, работ, услуг в Российской Федерации // журнал Инженерный вестник Дона, 2017, №1 [Электронный ресурс]. Режим доступа: <http://ivdon.ru/ru/magazine/archive/n1y2017/4006>, свободный. – Загл. с экрана. – Яз. рус.
6. Эконометрика: учебник для магистров / И.И. Елисеева (и др.); под ред. И.И. Елисеевой. М: Издательство Юрайт, 2012. 453 с.
7. Ворокова Н.Х. Методы и модели анализа и прогнозирования производительности труда / Н.Х. Ворокова, А.Е. Сенникова // Научное и образовательное пространство: перспективы развития: материалы II Междунар. науч.-практ. конф. (Чебоксары, 24 апр. 2016 г.) / редкол.: О.Н. Широков [и др.] – Чебоксары: ЦНС «Интерактив плюс», 2016. – С. 375–381.
8. Доугерти К. Введение в эконометрику. – М.: Финансы и статистика, 1999. – 215 с.
9. Афанасьев В.Н. Анализ временных рядов и прогнозирование: учебник / В.Н. Афанасьев, М.М. Юзбашев. – 2-е изд., перераб. И доп. – М.: Финансы и статистика; ИНФРА-М, 2012. – 320 с.
10. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа: пер. с нем. – М.: Финансы и статистика, 1982. – 305 с.

Shumilina Vera Evgenevna

Don state technical university, Rostov-on-Don, Russia
E-mail: Shumilina.vera@list.ru

Tsvil Maria Mikhaylovna

Russian customs academy
Rostov branch, Rostov-on-Don, Russia
E-mail: tsvilmm@mail.ru

Building a time series regression model with the aim of predicting the labor productivity index in the Russian Federation

Abstract. As part of the Economic Security Strategy of the Russian Federation until 2030, adopted by the Decree of the President of the Russian Federation in 2017, one of the monitored indicators is the labor productivity index, which is an important indicator of the economic efficiency of production, the realization of labor and human potential, the development of equipment and technology. The purpose of this study is to forecast the labor productivity index for the Russian Federation for subsequent periods. In the course of the study, the specificity of the relationship by the series of dynamics was demonstrated in drawing up the regression model. As the time series, the dynamics of investment in fixed assets (x_t) and changes in the labor productivity index (y_t) in the Russian Federation for the period 2008–2017 are considered. The study of the previously obtained regression equation $y(x)$ revealed the presence of autocorrelation in the residues ($y - \hat{y}_x$). It was found that these series of dynamics are characterized by the presence of a trend. The regression models of the x_t , y_t dynamics series considered during the study have polynomial trends of the fourth degree, improved with the help of dummy variables. When constructing a regression model for these time series, it became necessary to take into account trends. The method of accounting for trends was chosen – the method of deviations from the trend. In this case, the regression model is built on residual values. For this purpose, residual values d_y and d_x are found for each of the series. The study obtained a statistical model of the dependence of the labor productivity index on the dynamics of investment in fixed assets in the whole of the Russian Federation, which allows us to forecast labor productivity index for subsequent periods.

Keywords: labor productivity; forecasting; modeling; fixed capital investments; regression analysis; fourth degree polynomial; determination coefficient; autocorrelation; trend