УДК 336.02

Юров В.М.
Московский государственный областной университет

МОДЕЛИРОВАНИЕ НЕСТАЦИОНАРНЫХ ВРЕМЕННЫХ РЯДОВ С ВЫРАЖЕННЫМИ КОЛЕБАНИЯМИ С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТОВ EXCEL

Аннотация. Основным инструментом экономической науки в настоящее время являются временные ряды. В статье рассматривается компьютерная технология моделирования временных рядов с выраженными колебаниями с использованием инструментов EXCEL. Для моделирования используются модель тренда и модели авторегрессии и скользящего среднего. В качестве примера рассматривается временной ряд курса евро на ММВБ (июль–сентябрь 2014 г.) При моделировании использовались все необходимые статистические процедуры, требуемые для идентификации и оценки параметров модели и проверки ее адекватности и точности. На основе формальных критериев осуществляется выбор лучшей модели с целью получения наиболее точных прогнозных оценок.

Ключевые слова: временной ряд, авторегрессия, скользящее среднее, автокорреляционная функция, корреляционная матрица, значимость коэффициентов модели.

V. Yurov
Moscow State Regional University

MODELLING NON-STATIONARY TIME SERIES WITH STRONG OSCILLATIONS BY USING EXCEL

Abstract. The main modern instrument of economic science is time series. The article deals with computer simulation technology of time series with strong oscillations using the tools of Excel. The trend model, the model of auto-regression and moving average are used for the simulation. As an example we consider the time series of the euro on the MICEX from 07 to 09.2014. All necessary statistical procedures are used in the simulation to identify and estimate the parameters of the model and verify its adequacy and accuracy. Formal criteria are used to select the best model in order to get the most accurate assessment of prognosis.

Key words: time series, auto-regression, moving average, autocorrelation function, correlation matrix, the significance of the coefficients of the model.

Для анализа поведения временных рядов с выраженными колебаниями и построения математических моделей, описывающих это поведение, широко используют линейную стохастическую модель авторегрессии и скользящего среднего ARMA (или авторегрессии и проинтегрированного скользящего среднего ARIMA)[1] . Эта модель связывает текущее значение изучаемой переменной со значениями этой же переменной в предыдущие моменты

[1] Аббревиатура от англ. *autoregressive moving-average model* и *autoregressive integrated moving average*.

времени, а также с текущим и предыдущими значениями остатков модели. Данная модель чрезвычайно популярна и практика подтвердила ее точность и гибкость. Однако построение модели ARMA – сложный процесс. Его не так просто реализовать, и требуется большая практика, чтобы овладеть им. Эта модель используется во многих статистических пакетах. С целью применения этой модели в практической деятельности широким кругом экономистов, а также преподавателями и студентами в учебном процессе представляется целесообразным рассмотреть процесс построения модели ARMA в программе EXCEL.

Процесс построения модели включает три этапа: идентификацию модели, оценивание ее параметров и диагностику модели.

Идентификация модели осуществляется путем определения и анализа автокорреляционной и частной автокорреляционной функций (АКФ и ЧАКФ) ряда. Это возможно только для стационарных временных рядов. На практике экономические ряды, как правило, являются нестационарными. Однако во многих случаях их можно свести к стационарным временным рядам путем выделения тренда или с помощью перехода к рядам конечного числа разностей. Если после выделения тренда ряд стал стационарным, то для такого ряда строится ARMA модель вида

$$\hat{Y}_t = \sum_{i=1}^{p} a_i Y_{t-i} - \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t,$$

где Y_t, $\hat{Y}_t$ – фактическое и расчетное значения уровня стационарного ряда в момент времени t;

a_i – коэффициент уравнения авторегрессии i -го порядка;

b_j – коэффициент уравнения скользящего среднего j-го порядка;

ε_t – отклонение фактического уровня $\hat{Y}_t$ от расчетного Y_{t-i} уровня ряда в момент времени t.

Если после детрендирования ряд остался нестационарным, то его приводят к стационарному путем перехода к ряду разностей размерностью d. В этом случае строится ARIMA-модель (модель Бокса – Дженкинса) вида

$$\Delta Y_t^d = \sum_{i=1}^{p} \Delta Y_{t-i}^d - \sum_{j=1}^{q} b_j \varepsilon_{t-j} + \varepsilon_t,$$

где ΔY_t^d, $\Delta \hat{Y}_t^d$ – уровни преобразованного ряда, образованного разностями размерностью d фактических и расчетных значений уровней исходного ряда в момент времени t;

ε_t – отклонение фактического от расчетного $\hat{Y}_t^d$ уровня преобразованного ряда в момент времени t.

На этапе идентификации производится выбор некоторой частной модели из всего класса ARMA-моделей, т. е. выбор значений p, q и d. Для определения вида и порядка процессов, порождающих стационарный временной ряд используют аппарат автокорреляционных функций: АКФ и ЧАКФ. При этом в целом можно выделить три возможных ситуации [2]:

а. Процесс авторегрессии 1–3 порядка проявляет себя тем, что АКФ экспоненциально убывает либо представляет собой смесь синусоиды и убывающей экспоненты, а ЧАКФ имеет первые 1–3 значимых (ненулевых) значений, после чего становится неотличимой от ЧАКФ белого шума. Количество значимых значений ЧАКФ дает порядок авторегрессии p.

b. Процесс скользящего среднего 1–3 порядка проявляет себя обратной ситуацией, когда ЧАКФ экспоненциально убывает либо представляет собой смесь синусоиды и убывающей экспоненты, а АКФ наоборот имеет одно или несколько значимых значений, после чего становится неотличимой от АКФ белого шума. В этом случай количество ненулевых значений АКФ дает порядок скользящего среднего q.

c. Смесь двух процессов приводит к появлению экспонент и синусоид в обеих функциях, однако если имеются две чистые синусоиды, то это свидетельствует о наличии процесса первого порядка и для авторегрессии, и для скользящего среднего. В целом количество ненулевых значений АКФ и ЧАКФ дает порядок q и p.

Другие формы АКФ и ЧАКФ свидетельствуют о том, что во временном ряде имеется тренд (практически неубывающая на отрезке 15–20 значений лага АКФ), либо сезонные колебания (АКФ в виде практически неубывающей синусоиды), от которых следует избавится, для того чтобы приступить к идентификации внутренних параметров ряда.

Выбор порядка d модели ARIMA основан на анализе поведения АКФ рядов Y_t, ΔY_t^1, ΔY_t^2. Требуемый порядок достигнут, 1если АКФ быстро затухает.

Однако рассмотренный способ идентификации модели может использоваться, если временной ряд является достаточно длинным (с числом уровней более 50) [2]. В противном случае он может при последующем анализе привести к выводу о непригодности идентифицированной модели и необходимости замены ее альтернативной моделью.

При использовании современных компьютеров наиболее действенным путем решения этой задачи является перебор всех возможных моделей. На практике в большинстве случаев порядок параметров p, d, q не превышает двух [2], поэтому число возможных моделей невелико.

На втором этапе производится ***определение оценок коэффициентов модели***. Для этого используется метод наименьших квадратов. При этом минимизируется сумма квадратов отклонений расчетных уровней ряда от $(\hat{Y}_t, \Delta\hat{Y}_t^d)$ фактических уровней : $(Y_t, \Delta Y_t^d)$

$$\sum_{t=2}^{n}(Y_t - \hat{Y}_t)^2 \longrightarrow min;$$

$$\sum_{t=2}^{n}(\Delta Y_t^d - \Delta\hat{Y}_t^d)^2 \longrightarrow min;$$

при соблюдении ограничений на коэффициенты исходя из условий стационарности AR-части модели и обратимости MA-части. Метод наименьших квадратов будем реализовывать так называемым эволюционным численным методом надстройки EXCEL «Поиск решения». Этот метод позволяет определять глобальный минимум функции при наличии нескольких экстремумов.

Если модель имеет только AR-часть, то для определения оценок коэффициентов целесообразнее использовать функцию «Регрессия» надстройки «Анализ данных». Для оцененных коэффициентов вычисляются прибли-

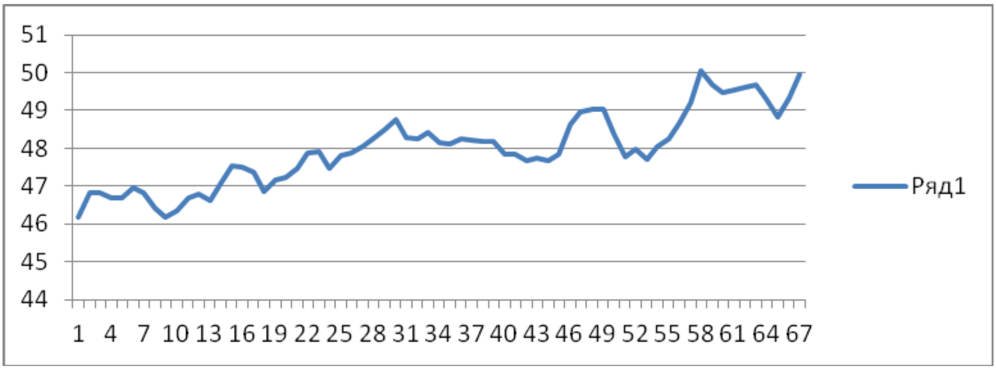

Рис. 1. Изменение курса евро в июле-сент. 2014 г.

женные стандартные ошибки, дающие возможность, при дополнительном предположении о нормальности распределения величин Y_{t-i} и ε_{t-i}, строить доверительные интервалы для этих коэффициентов и проверять гипотезы об их истинных значениях с целью уточнения спецификации модели с помощью критерия Стьюдента.

На третьем этапе применяются различные ***диагностические процедуры для проверки адекватности выбранной модели имеющимся данным***. Неадекватности, обнаруженные в процессе такой проверки, могут указать на необходимую корректировку модели, после чего производится новый цикл подбора, и т. д. до тех пор, пока не будет получена удовлетворительная модель.

Рассмотрим компьютерную технологию выполнения расчетов при построении модели авторегрессии и скользящего среднего инструментами EXCEL. Методику построения моделей рассмотрим на примере временного ряда курса евро, имеющего 67 уровней с 1 июля по 30 сентября 2014 г. График этого ряда представлен на рис. 1.

Визуальный анализ графика ряда показывает, что ряд имеет выраженные колебания уровней ряда и линейный тренд, поэтому ряд предположительно является нестационарными. Для более обоснованного вывода о нестационарности ряда определим АКФ ряда. Обычно достаточно определить первые 10–15 значений коэффициентов автокорреляции. Для этого на рабочем листе EXCEL необходимо в столбцах разместить значения уровней ряда со сдвигом каждого столбца на одну строку относительно предыдущего и далее применить функцию надстройки «Анализ данных» «Корреляция». Рассчитанные значения коэффициентов автококорреляции представлены в табл. 1. С увеличением лага отсутствует их быстрое затухание, что говорит о нестационарности ряда.

Таблица 1

Коэффициенты автокорреляции ряда «курс евро»

Лаг	1	2	3	4	5	6	7	8	9	10
АКФ	0,923	0,825	0,758	0,691	0,639	0,613	0,612	0,658	0,697	0,715

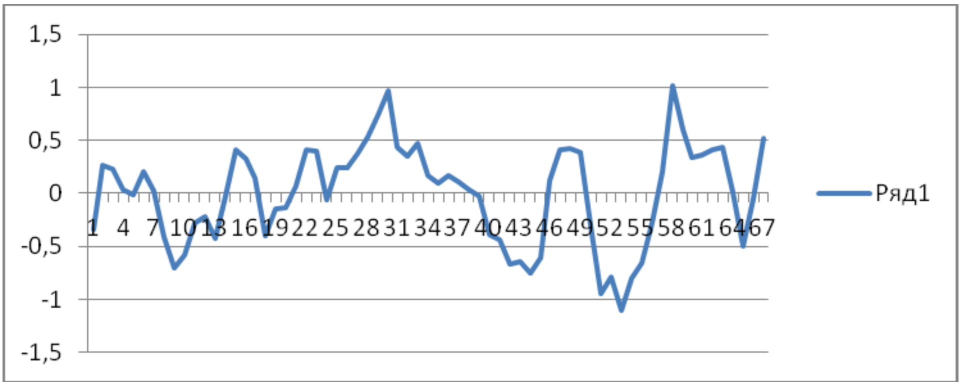

Рис. 2. Детрендированный ряд «курс евро»

С помощью функции «Регрессия» находим уравнение линейного тренда

$$= 46{,}48+0{,}04413*t.$$

Коэффициенты этого уравнения значимы с P значениями, существенно меньшими 0,05. Показатели точности этой модели: вариация остатков $SS_{ост}$ = 14,32, стандартная ошибка E=3,784; процентная ошибка $E\%$=7,82%. Исключив тренд из уровней исходного ряда, получаем детрендированный ряд. График этого ряда представлен на рис. 2. АКФ и ЧАКФ этого ряда представлены в табл. 2.

Таблица 2

Значения АКФ- и ЧАКФ-рядов «курс евро»

Лаг	1	2	3	4	5	6	7	8	9	10	11	12
АКФ	0,764	0,457	0,225	0,016	-0,132	-0,200	-0,193	-0,033	0,103	0,150	0,099	0,014
ЧАКФ	0,759	-0,359	0,013	-0,105	-0,057	-0,037	0,089	0,265	-0,083	-0,072	-0,044	-0,165

Границы, при выходе из которых коэффициенты являются значимыми с вероятностью, близкой к 0,95 (P значение близко к 0,05), определяются по формуле [5]:

$$\pm\frac{2}{\sqrt{n}}=\pm\frac{2}{\sqrt{67}}=\pm 0{,}244$$

где n – число уровней ряда. Из этого диапазона выходят по два первых значения АКФ и ЧАКФ, остальные значения образуют синусоиду с убывающей амплитудой, следовательно, детрендированный ряд является стационарным. Данный ряд можно описать моделью ARMA с числом членов авторегрессии p и скользящего среднего q не больше двух.

Проведем теперь оценивание параметров модели. Для применения метода наименьших квадратов получим выражение для вычисления остатков ε_t

$$\varepsilon_t=Y_t-\hat{Y}_t=Y_t-\sum_{i=1}^{p} a_i Y_{t-i}+\sum_{j=1}^{q} b_j\varepsilon_{t-j},$$

В этом выражении отсутствуют несколько начальных значений Y_{t-i} и ε_{t-i}. Поскольку рассматриваемый ряд является достаточно длинным (n>50), то в соответствии с [2] используем приближение, заключающееся в приравнивании q начальных значений ε_{t-j} и p значений Y_{t-i} их математическим ожиданиям, т. е. 0. Для длинного ряда такая потеря информации несущественна. Исходные данные и расчетные показатели для первых 10 уровней при p=1 и q=1 приведены в табл. 3.

Таблица 3

Исходные данные и расчетные показатели уровней и остатков ряда

t	Y_t	$\hat{Y}_t$	ε_t	ε_t^2
1	-0,33943	0	0	0
2	0,26724	-0,20334	0,470575	0,22144129
3	0,22941	0,360719	-0,13131	0,01724218
4	0,03178	0,081444	-0,04966	0,00246655
5	-0,01165	-0,00214	-0,00951	9,0506E-05
6	0,20202	-0,01103	0,213055	0,04539243
7	0,02529	0,211856	-0,18657	0,03480681
8	-0,41644	-0,06439	-0,35205	0,12393766
9	-0,71027	-0,39956	-0,31071	0,09653849
10	-0,5865	-0,55796	-0,02854	0,00081473

Для определения параметров модели b_j используем надстройку EXCEL «Поиск решения». Целевой функцией при этом будет вариация остатков $SS_{ост} = \sum \varepsilon_t^2$ Изменяемые ячейки – ячейки с начальными значениями параметров. Ограничения задачи - условия стационарности для AR части и обратимости MA части модели [1]:

при p=1 $|\alpha_1| \leq 1$;

при p=2 $|\alpha_1| < 2$, $a_2 < 1 - |\alpha_1|$;

при q=1 $|b_1| \leq 1$;

при q=2 $|b_1| < 2$, $b_2 < 1 - |b_1|$.

Минимальные значения вариация остатков $SS_{ост}$ для разных сочетаний p и q представлены в табл. 4.

Таблица 4

Минимальные значения вариации остатков

Вариация остатков			
$p \backslash q$	0	1	2
0		6,967	5,556
1	6,124	5,374	5,307
2	5,471	5,374	5,189

Наименьшее значение $SS_{ост}$ имеет самая сложная модель при p=2 и q=2. Однако следует иметь в виду, что более сложные модели обладают лучшими аппроксимирующими свойствами и одновременно менее устойчивы, что для целей прогнозирования является определяющим отрицательным фактором. Поэтому для выбора лучшей модели используют так называемые информационные критерии: Акаике, Шварца и др. Согласно критерию Акаике [3], среди альтернативных моделей выбирается та модель, для которой минимизируется величина

$$AIC= \ln [SS_{ост} / (n-p-q)]+ 2 (p+q)/n.$$

Здесь второе слагаемое представляет собой штраф за усложнение модели. Рассчитанные значения критерия AIC представлены в табл. 5.

Таблица 5

Расчетные значения критерия AIK

$p \backslash q$	0	1	2
0		-2,219	-2,415
1	-2,347	-2,448	-2,431
2	-2,431	-2,419	-2,424

Таким образом, выбираем модель с p=q=1. Коэффициенты этой модели *a1*=0,599; *b1*=-0,426. ARMA модель будет иметь вид

$$\hat{Y}_t = 0{,}599\,Y_{t-1} + 0{,}426\,\varepsilon_{t-1} + \varepsilon_t$$

Аналогичный результат получается и при использовании пакета STATISTICA. Проверим теперь значимость коэффициентов модели. Стандартные ошибки коэффициентов находятся по формуле [2]:

=0,1083

t – статистики для коэффициентов a_1 и b_1 находим по формулам:

$$t_{b1} = \frac{b_1}{\sigma_{b1}} = \frac{0{,}426}{0{,}1083} = 3{,}94$$

Уровень значимости P коэффициентов находим с помощью функции EXCEL «СТЬЮДРАСП» (t; n-k). При числе степеней свободы n-k=67-2=65 находим: $\boldsymbol{P}_{a1}$=3,055E-07; $\boldsymbol{P}_{b1}$ =0,00108. Таким образом, коэффициенты модели значимы. График остатков модели ε_t представлен на рис. 3.

Диагностика оцененной модели направлена на проверку гипотезы о том, что остатки модели ε_t образуют процесс белого шума, т. хе. имеют нулевую автокорреляцию, и подчинены нормальному закону распределения. При проверке значимости АКФ-остатков используют два подхода [5]:

проверка значимости каждого коэффициента автокорреляции отдельно;

проверка значимости группы коэффициентов автокорреляции с помощью теста Бокса-Льюнга.

В табл. 6 приведены результаты расчетов коэффициентов автокорреляции остатков r_k, их стандартные ошибки $\sigma(r_k)$, статистика Бокса-Льюнга Q и уровень значимости для этой статистики P. Коэффициенты автокорреляции определялись с помощью функции EXCEL КОРРЕЛЯ-

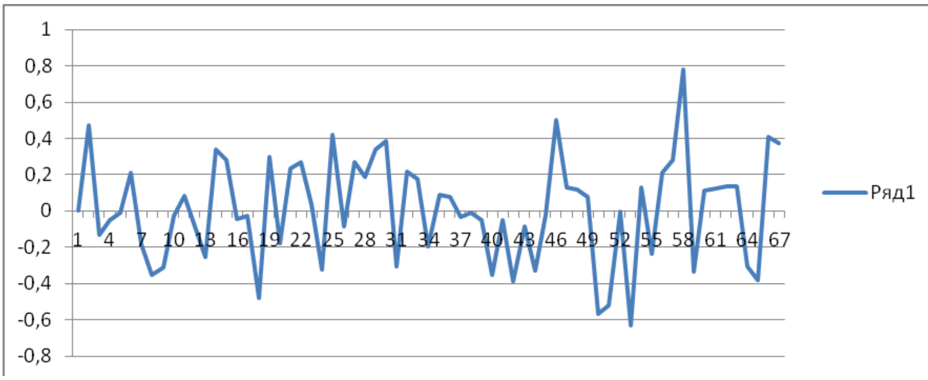

Рис. 3. График остатков модели

ЦИЯ, стандартные ошибки определялись в соответствии с [4] как:

$$\sigma(r_k) = \sqrt{\frac{\left(\frac{1}{n}\right)(n-k)}{(n+2)}}$$

где n – число наблюдений. Если r_k меньше по модулю величины $2\sigma(r_k)$, то коэффициенты автокорреляции r_k являются статистически незначимы с вероятностью 0,95.

Статистика Q на данном лаге κ определялась как $Q_k=n(n+2)\sum_{i=1}^{k}\frac{r_i^2}{n-i}$. При выполнении нулевой гипотезы отсутствия автокорреляции Q-статистика имеет распределение $\chi^2(\kappa\text{-}p\text{-}q)$. Уровни значимости P_κ, соответствующие статистике Q_κ, можно определить с помощью функции Excel ХИ2РАСП(Q_κ, κ). Если P_κ больше заданного уровня значимости, то κ первых значений автокорреляционной функции остатков статистически незначимы.

Таблица 6

Результаты расчетов по проверке значимости АКФ-остатков

Лаг k	r_k	$2\sigma(r_k)$	Q	P_κ
1	0,015186	0,238968	0,016154546	0,898861
2	0,111807	0,237151	0,905248066	0,635957
3	-0,02732	0,23532	0,959176834	0,811129
4	-0,09511	0,233474	1,623026684	0,804648
5	-0,11524	0,231614	2,61333798	0,759338
6	0,014462	0,229738	2,629189538	0,853738
7	-0,26933	0,227847	8,218302963	0,313738
8	0,082381	0,22594	8,750081895	0,363816
9	0,124202	0,224018	9,979654918	0,352135
10	0,104955	0,222078	10,87307949	0,367491
11	0,243984	0,220121	15,78736393	0,14921
12	-0,13569	0,218147	17,33498357	0,137425
13	0,01462	0,216155	17,35328163	0,183653
14	-0,06804	0,214144	17,75703471	0,218072
15	-0,09571	0,212114	18,57134685	0,233819

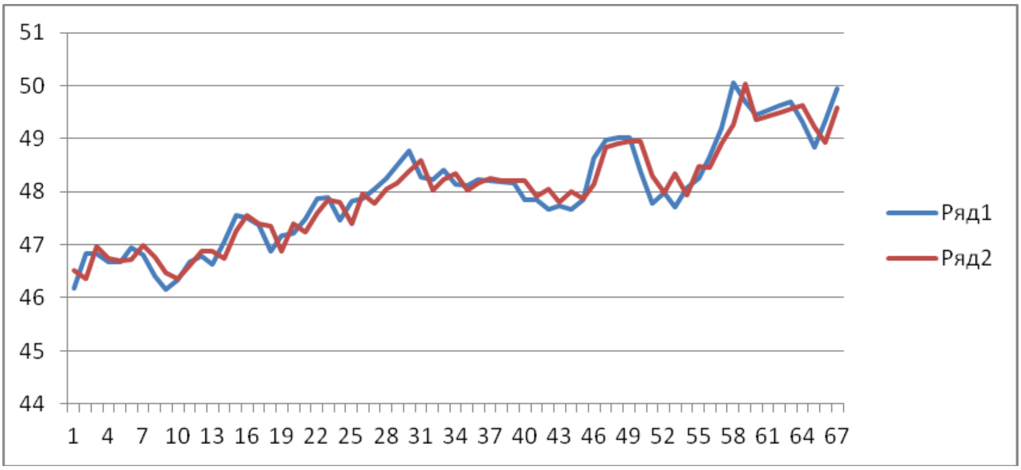

Рис. 4. Сравнение исходных и модельных значений уровней ряда «курс евро»

Из сопоставления полученных значений первых двух столбцов результатов следует, что коэффициенты корреляции остатков незначимы с вероятностью, близкой к 0,95 на всех 15 лагах. Из рассмотрения полученных значений последнего столбца таблицы также следует, что все *к* (*к*=1-15) первых значений АКФ остатков статистически незначимы. Таким образом, гипотеза о том, что остатки модели образуют процесс белого шума подтверждена.

Для проверки нормальности остатков существует множество тестов. Рассмотрим тест Харке-Бера [3]. Он вычисляет выборочные значения для коэффициентов асимметрии

$$s=\sum \frac{(\varepsilon_t-\bar{\varepsilon})^3}{n\sigma^3}$$

и эксцесса

$$\kappa=\sum \frac{(\varepsilon_t-\bar{\varepsilon})^4}{n\sigma^4}$$

При условии нормальности остатков статистика Харке-Бера $[S^2 + (K-3)^2/4](n-p-q-1)/6$ имеет χ^2 распределение с двумя степенями свободы. Рассчитанные значения: S=0,0560; K=2,75; статистика Харке-Бера=0,198; уровень значимости P=0,906, т. е. нулевая гипотеза о нормальности распределения остатков не отвергается. Таким образом, условия адекватности модели выполнены.

С учетом модели тренда построенная модель временного ряда имеет вид $\hat{Y}_t = 46{,}48+0{,}04413*t+0{,}599Y_{t-1} +0{,}426\varepsilon_{t-1}+\varepsilon_t$.

Показатели точности этой модели: вариация остатков $SS_{ост}$ = 5,373; стандартная ошибка E=2,318; процентная ошибка E %=4,79 %, т.е. по сравнению с моделью тренда ошибка модели меньше на 38,9 %.

В заключение (рис. 4) представим график исходных и модельных значений уровней рассматриваемого временного ряда. Представленный график свидетельствует о высокой точности построенной модели и ее пригодности для прогнозирования.

Рассмотренная технология моделирования временного ряда в среде EXCEL позволила получить модель ряда, аналогичную моделям, полученным с использованием профессиональных статистических пакетов. При этом использовались все необходимые статистические процедуры, необходимые для идентификации и оценки параметров модели и проверки ее адекватности и точности. В отличии от статистических пакетов, моделирование в EXCEL доступно более широкому кругу специалистов и студентов, наглядно представляет процесс моделирования и дает широкую свободу для применения различных статистических инструментов, критериев и тестов для улучшения качества модели.

ЛИТЕРАТУРА:

1. Айвазян С.А. Методы эконометрики: учебник. М.: Инфра-М, 2014. 512 с.
2. Бокс Дж., Дженкинс Г. Анализ временных рядов (прогноз и управление): в 2-х вып. Вып. 1. М.: Мир, 1974. 408 с.
3. Канторович Г.Г. Анализ временных рядов // Экономический журнал ВШЭ. 2002. № 1. С. 85–116.
4. Новиков А.И. Эконометрика: учеб. пособ. для бакалавров. М.: Дашков и К^0, 2012. 224 с.
5. Носко В.П. Эконометрика: в 2-х кн. Кн. 1 (Основные понятия, элементарные методы; регрессионный анализ временных рядов). М.: Дело, 2011. 672 с.
6. Протасов Ю.М. Моделирование взаимосвязей макроэконо-мических показателей средствами корреляционно-регрессионного анализа // Вестник МГОУ (Электронный журнал). 2012. № 2. С. 177–181.