

УДК 517.534

А.А. ЧИСТЯКОВА, Б.В. ШАМША

Харьковский национальный университет радиоэлектроники, Украина

ИДЕНТИФИКАЦИЯ СТРУКТУРЫ НЕСТАЦИОНАРНОГО ВРЕМЕННОГО РЯДА ПРИ ПОМОЩИ МЕТОДА СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА

В работе рассмотрен метод сингулярного спектрального анализа (SSA) для идентификации структуры нестационарных временных рядов. Целью метода является выделение компонент временного ряда, таких как тренд и периодическая составляющая. Решение данной задачи необходимо для построения модели временного ряда и определения завуалированных зависимостей. Проведен анализ структуры нестационарного временного ряда цен на сахар. Даны рекомендации по выбору параметров при использовании метода SSA для рядов, которые не могут быть приведены к однородным. Построена модель нестационарного временного ряда с учетом компонент тренда и периодики.

Ключевые слова: нестационарные временные ряды, идентификация модели, сингулярный спектральный анализ, главные компоненты.

Введение

На сегодняшний день информационные технологии охватывают практически все отрасли производства, науки, где требуется обработка информации. Современные информационные технологии (ИТ) – важнейший источник развития различных сфер жизнедеятельности и являются самым эффективным средством хранения, обработки и выдачи информации. В связи с ростом хранимых, упорядоченных во времени, данных о характеристиках объектов, процессов и систем в промышленности, экономике, медицине, образовании, социологии расширяются возможности анализа и прогнозирования количественных и качественных изменений этих характеристик, а также их носителей. Систематическое и комплексное исследование тенденций развития процессов на основе анализа временных рядов (ВР), извлеченных из хранилищ и баз данных, является сферой профессиональной деятельности специалистов различного профиля: менеджеров среднего звена, аудиторов, специалистов в области контроля качества, экономистов, руководителей, маркетологов, аналитиков и др. Корректное моделирование и прогнозирование временных рядов сложной структуры дает возможность уменьшить риски при планировании, администрировании, а также принятии управленческих решений.

Постановка задачи в общем виде

Множество процессов в современном мире являются хаотическими, случайными, нестационарными и тем самым сложно предсказуемыми, однако, при корректной идентификации модели данных

процессов, можно добиться качественных результатов. Обычно для моделирования нестационарных временных рядов требуется приведение к стационарности, используя следующие приемы:

- выделение линейного тренда или его удаление;
- взятие разностей (так, например, преобразование взятия разности первого порядка позволяет избавиться от линейного тренда в ряде; смысл заключается в том, что из текущего значения ряда вычитается предыдущее со сдвигом 1 и результат предоставляется в качестве значения нового ряда);
- удаление автокорреляций;
- вычисление остатков;
- логарифмическое преобразование.

Для стационарных временных рядов уже разработано достаточное количество методов их моделирования и прогнозирования, выбор которых зависит от предпочтений и выводов исследователя.

Однако, существуют такие временные ряды, которые не могут быть приведены к стационарным, в данных рядах отсутствует точка равновесия, они подвержены большей дисперсии – изменениям, вызываемым влиянием возмущений или внешних переменных. Выбор структуры для таких временных рядов является сложной и неоднозначной задачей. В этом случае целесообразно использовать сингулярное разложение, корректно выбирая его параметры. Структура нестационарных временных рядов достаточно сложна, и ее определение является основным этапом для построения модели ряда и возможности его прогнозирования. Данные временные ряды описывают сложные процессы и часто встречаются в современном мире и различных отраслях науки:

социологии, медицине, экономике и т. д. Статистические методы оценки нестационарных временных рядов позволяют получить завуалированные зависимости между величинами, которые дают возможность идентифицировать модель прогнозирования. Результаты прогнозирования являются основой для принятия управленческих решений, поэтому идентификация модели имеет большое практическое значение.

В настоящее время основными статистическими методами исследования временных рядов являются: метод выделения тренда (временного сглаживания), регрессионный, автокорреляционный, адаптивный (скользящих средних), метод гармонического анализа, сингулярного спектрального анализа, метод Бокса-Дженкинса (ARIMA), бутстрепа (численного размножения выборок) и нейросетевой. Однако все вышеперечисленные методы можно применять только для стационарных временных рядов.

Для выделения структуры нестационарного временного ряда в данной работе предлагается использовать метод, основанный на сингулярном разложении, который в иностранных изданиях больше известен как SSA (Singular Spectrum Analysis). В работе предлагается проверить адекватность модели нестационарного временного ряда, построенного при помощи данного метода. Также предлагается исследовать данный метод с точки зрения выбора наиболее эффективных параметров на примере нестационарных временных рядов.

Анализ последних исследований и публикаций

В работе для идентификации модели нестационарного временного ряда был выбран метод Singular Spectrum Analysis (SSA) или Анализ сингулярного спектра (ACC), который зародился в 70х-80х годах прошлого столетия и вызвал особый интерес исследователей. Ссылки на основную литературу по данному методу можно найти в литературе [1-3]. Данный метод имеет строгое обоснование в рамках теории динамически систем и используются главным образом для определения основных составляющих временного ряда, а также подавления шума. За время своего существования метод значительно расширился, а также появились примеры его применения в различных областях: медицине, социологии, экономике пр. Так как SSA базируется на эмпирической ортогональной функции, то появляются возможности прогнозировать временные ряды, правильно используя данный метод.

В основе SSA для анализа временных рядов $\{y_t, y_{t+1}, y_{t+2}, \dots, y_n\}$ лежит построение множества

векторов задержек:

$$Y_t = (y_t \ y_{t-1} \ \dots \ y_{t-L+1})^T, \quad t=L, L+1, \dots, N.$$

Метод задержек, посредством которого строятся указанные векторы, устанавливает переход от исходного одномерного (скалярного) временного ряда к многомерному (векторному) представлению, похожему к применяемому в авторегрессии. При этом каждый вектор строится из некоторого числа L следующих друг за другом значений исходного временного ряда. Данные вектора образуют траекторную матрицу, которая впоследствии подлежит обработке по алгоритму, близкому к методу главных компонент [4]. Суть метода главных компонент лежит в снижении размерности сходного пространства факторов (в данном случае пространство векторов) посредством перехода к более информативным переменным. Полученные таким образом новые переменные называются главными компонентами, которые дают возможность отделить основные тенденции временного ряда, выделить его структуру, а также отсеять случайные возмущения.

Таким образом, с помощью метода SSA возможно определить завуалированную структуру временного ряда, выбрав правильно параметры метода и главные компоненты.

Постановка задачи исследования

Главной идеей сингулярного спектрального анализа является преобразование одномерного ряда в многомерный с помощью однопараметрической сдвиговой процедуры и исследование полученной многомерной траектории с помощью анализа главных компонент. Известно, что обобщенную модель временного ряда можно представить в виде:

$$Y(t) = \sum_{i=1}^n T(t) + \sum_{j=1}^l G(t) + A(t), \quad (1)$$

где $\sum_{i=1}^n T(t)$ – сумма простейших трендовых составляющих, соответствующая основным векторам временного ряда;

$\sum_{j=1}^l G(t)$ – сумма элементарных гармонических составляющих, определяющаяся совокупностью различных периодических факторов;

$A(t)$ – случайная составляющая, образованная совокупностью случайных факторов.

Таким образом, используя метод SSA возможно выделить составляющие временного ряда, что является предпосылкой к выбору вида модели временного ряда и метода прогнозирования.

Так как нестационарные временные ряды обладают свойствами хаотичности, кластеризации вола-

тельности, то им присущи скрытые компоненты [5], выделение которых зависит от правильного выбора параметров в методе SSA. Целью данной работы является исследование выбора параметров метода сингулярного спектрального анализа для нестационарных временных рядов.

Идентификация модели нестационарного временного ряда при помощи метода SSA

В качестве примера нестационарного временного ряда будем использовать динамику изменения цены на сахар в долларах за фунт в период с 02.01.2002 г. по 22.07.2009 г. Длина данного временного ряда составляет 1903 значения, график данных приведен на рис. 1.

После первичного исследования данного ряда было доказано, что ряд невозможно привести к стационарному, поэтому стандартные методы моделирования и выделения структуры данного ряда не смогут отразить все имеющиеся компоненты. В данном случае достаточно для определения структуры будем применять метод SSA.

Первым шагом метода для выделения компонент временного ряда $\{y_t, y_{t+1}, y_{t+2}, \dots, y_n\}$ является построение траекторной матрицы ряда, которая имеет вид матрицы векторов определенной длины L (длины выборки), $L, L \in \mathbb{N}, 1 < L < N$ и составляется матрица развертки Y , содержащая $K=N-L+1$ векторов развертки, имеющих заданную размерность L . То есть матрица имеет вид:

$$Y = \begin{pmatrix} y(t)_0 & y(t)_1 & \dots & y(t)_{N-L} \\ y(t)_1 & y(t)_2 & \dots & y(t)_{N-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ y(t)_{L-1} & y(t)_L & \dots & y(t)_{N-1} \end{pmatrix}, \quad (2)$$

где $y(t)_i$ - определенное i -тое значение исходного временного ряда;

Y - матрица развертки.

До сих пор не существует единственного обоснованного метода выбора L , поэтому данный параметр определяется эмпирически, хотя он, несомненно, имеет достаточно большое влияние на качество итоговой модели.

В данной работе проводится анализ выбора параметра L для идентификации структуры нестационарного временного ряда.

Следующим этапом метода является сингулярное разложение полученной матрицы, результатом которого является представление временного ряда в виде суммы:

$$Y = Y_1 + Y_2 + \dots + Y_d, \quad Y_k = \sqrt{\lambda_k} U_k V_k^T, \quad (3)$$

где U_k - k -й левый сингулярный собственный вектор; V_k - k -й правый сингулярный вектор или факторный вектор; λ_k - k -е сингулярное число.

Набор $(\sqrt{\lambda_k} U_k V_k^T)$ называют k -й собственной тройкой (eigentriple).

На третьем этапе, исходя из предположения, что каждое сингулярное число характеризует определенную составляющую нестационарного временного ряда, можно определить степень влияния каждого сингулярного числа на результирующий ряд и выделить составляющие процесса. При этом следует учитывать, что определение и группировка элементов сингулярного разложения для выделения таких компонент как тренд, периодическая составляющая также имеют большое влияние на качество итоговой модели. Данный процесс можно отнести к параметрам метода SSA, который выбирается исследователем на основе графического представления данных.



Рис. 1. Временной ряд изменения цены на сахар

Следующим этапом идентификации структуры ряда является процесс восстановления по выбранным главным компонентам. С геометрической точки зрения операция получения главных компонент есть изображение исходной выборки в базисе, составленном из выбранных собственных векторов, а восстановление – проектирование исходной выборки на гиперплоскость, порожденную выбранным набором собственных векторов ковариационной матрицы.

Итак, в методе SSA было выделено два параметра. Первый – это целое число L , длина каждого вектора траекторной матрицы, а второй параметр является структурным – это способ группировки главных компонент.

Дадим несколько рекомендаций по выбору длины гусеницы:

- чем больше длина векторов, образующих матрицу траекторного пространства, тем более детально получается разложение исходного ряда. Таким образом, наиболее детальное разложение достигается при выборе данного, приблизительно равного половине длины ряда ($L \approx N / 2$);

- сингулярные разложения одного и того же ряда длины N , соответствующие выбору длины гусеницы L и $N-L+1$ эквивалентны. Следовательно, для анализа структуры временного ряда не имеет смысла брать длину гусеницы, большую чем половина длины ряда;

- маленькая длина гусеницы может привести к смешиванию интерпретируемых компонент ряда, следовательно, не все главные компоненты будут выделены «в чистом виде»;

- при решении задачи выделения периодической компоненты с периодом τ следует выбирать длину гусеницы L кратной τ , то есть необходимо учитывать явную гармонику исходных данных;

- в общем, метод SSA устойчив относительно изменения длины векторов траекторной матрицы.

Также выделим важные моменты, связанные с отбором главных компонент. Пусть длина гусеницы L фиксирована и мы уже имеем сингулярное разложение траекторной матрицы исходного ряда. Тогда следующим шагом является группировка членов сингулярного разложения:

- если компонента ряда восстановлена только с помощью одной собственной тройки (собственное значение, собственный вектор и главная компонента) и оба сингулярных вектора имеют похожую форму, то восстановленная компонента будет иметь примерно такую же форму. Данное утверждение говорит о том, что после сингулярного разложения, исследуя графики собственных векторов, можно понять характер тренда. Например, если оба сингулярных вектора собственной тройки

похожи на линейные ряды, то соответствующая составляющая ряда также будет близкой к линейной. Если сингулярные векторы имеют экспоненциальную форму, то и компонента ряда будет такой же. Монотонные сингулярные векторы соответствуют монотонной компоненте ряда. Синусоидальные векторы порождают гармоническую составляющую ряда;

- чем больше собственное значение главной компоненты, тем больше вклад соответствующей восстановленной компоненты ряда.

Однако, стоит заметить, что лишь совместное корректное использования данных параметров дает лучший результат.

Итак, протестируем метод на предмет корректной идентификации параметров для выделения структуры нестационарных временных рядов на примере ряда, отражающего динамику роста цен на сахар (рис. 1).

После многочисленных экспериментов и исследованию динамических показателей данного ряда было принято решение выбрать длину векторов траекторной матрицы равную 1280. Данное значение больше половины длины ряда, также данная величина учитывает прохождения двух экстремумов временного ряда, то есть «память» каждого из векторов даст возможность учесть переломные моменты тренда в модели временного ряда. Итак после построения траекторной матрицы и ее разложения получили графики собственных векторов, представленные на рис. 2.

Так как тренд отражает плавно меняющуюся составляющую временного ряда, то при выборе компонент тренда основываемся на утверждении, что компонента должна содержать не больше экстремумов, чем исходный временной ряд. В данном случае выбираем компоненты 1-3.

Для анализа и выделения периодической компоненты воспользуемся двумерными графиками собственных векторов, которые представлены на рис. 3.

При выборе компонент, соответствующих периодике остановимся на тех, двумерные графики которых образуют окружность, т.е. вектора «закручивают» друг друга, показывая периодичность, отвечающую определенному периоду. В данном случае разные вектора выделяют разную периодичность, потому можно утверждать, что среди этих компонент есть сезонная, циклическая составляющие и т. п.

Итак, восстановление гармонической составляющей будем производить на основе 8-9, 13-14, 16-17 компонент.

Графики восстановленных компонент временного ряда показаны на рис. 4.

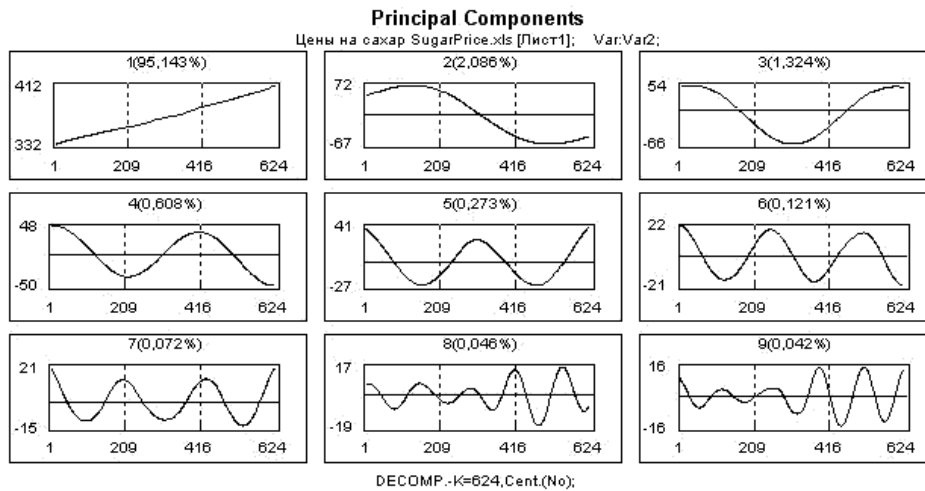


Рис. 2. Графики собственных векторов траекторной матрицы

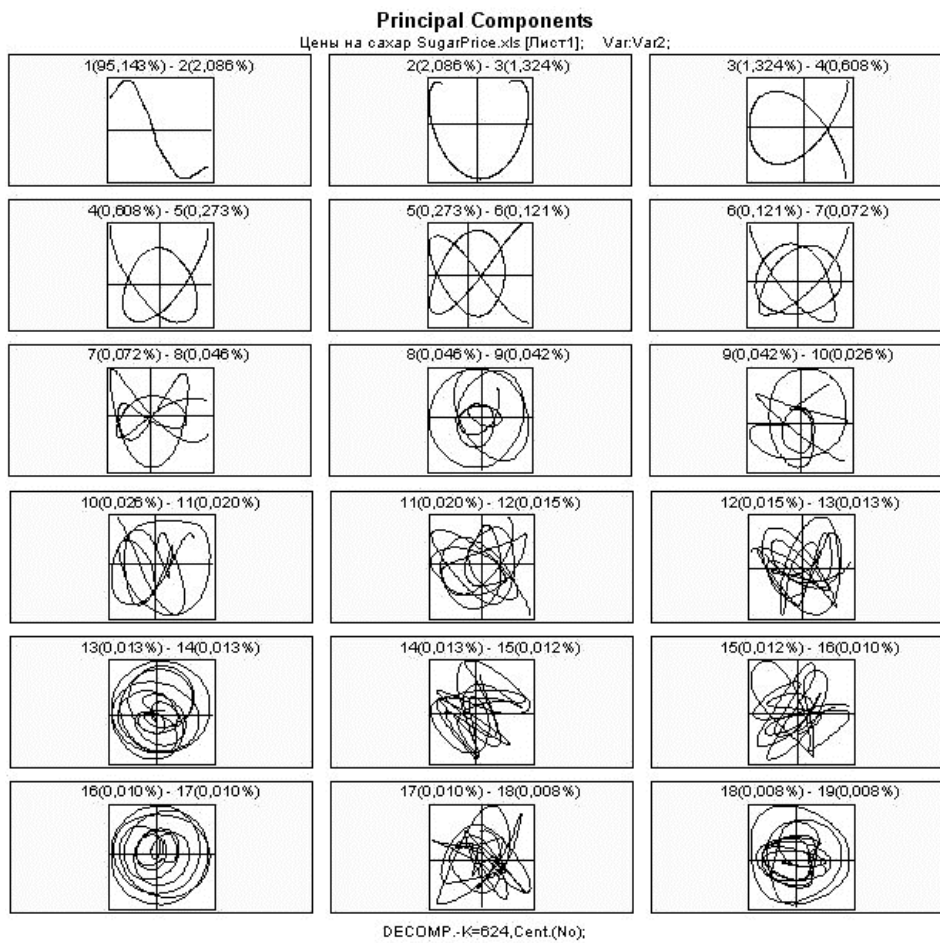


Рис. 3. Двумерные графики собственных векторов траекторной матрицы

Для возвращения к исходному ряду необходимо провести операцию усреднения по второстепенным диагоналям полученной траекторной матрицы Y . Для формирования каждой простейшей составляющей, определяемой соответствующим сингулярным числом, необходимо провести операцию усреднения по второстепенным диагоналям обратной матрицы развертки A_i^0 и получить ряд F_i .

$$\left. \begin{aligned}
 A_i^0{}_{L \times K} &= \sqrt{\lambda}_{L \times M} \times U_{i M \times i} \times V_{i \times K}^T \rightarrow F \\
 A_{(i-1)}^0{}_{L \times K} &= W_{L \times M} \times U_{(i-1) M \times (i-1)} \times V_{(i-1) \times K}^T \rightarrow F_{(i-1)}
 \end{aligned} \right\} \rightarrow \\
 \rightarrow F_i^0 = F_i - F_{i-1}. \tag{4}$$

Итак, после восстановления исходного ряда, по выбранным выше компонентам, получили результаты отраженные на рис. 5,6.

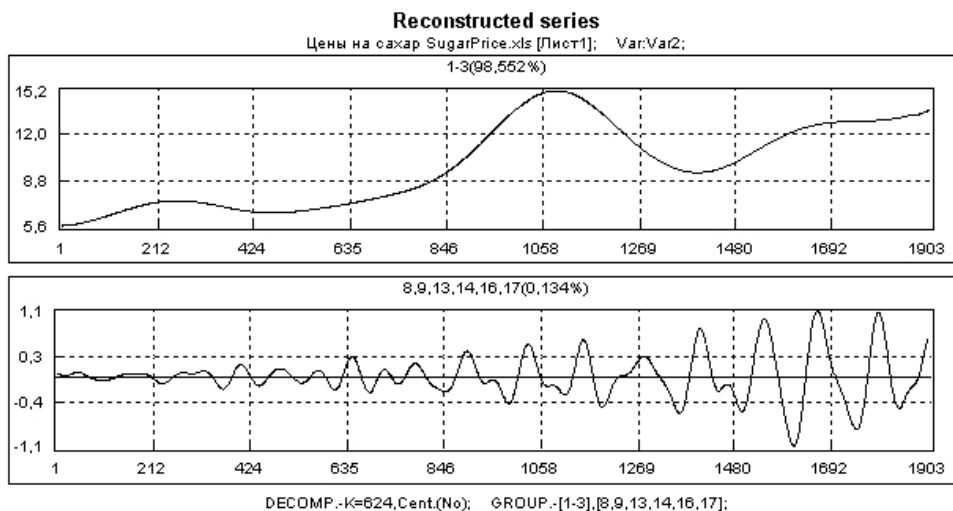


Рис. 4. Выделенные компоненты тренда и периодики исходного временного ряда



Рис. 5. Восстановленный временной ряд



Рис. 6. Погрешность модели нестационарного временного ряда

Заключение

В результате проведенного исследования можно утверждать, что возможно идентифицировать структуру нестационарных временных рядов, используя сингулярное разложение. Показано, что по фазовому пространству собственных чисел траекторной матрицы можно оценить наличие состав-

ляющих временного ряда. При правильном выборе параметра длины векторов, образующих траекторную матрицу, а также выделении компонент, соответствующих определенным признакам, метод анализа сингулярного спектра дает хорошие результаты. В дальнейшем, моделируя каждую из компонент ряда, можно строить модели прогнозирования исходного временного ряда.

При идентификации структуры нестационарного временного ряда были выделены гармонические компоненты с различными периодами, а также компонента нелинейного тренда.

Литература

1. Golyandina, N. *Analysis of Time Series Structure: SSA and Related Techniques [Text]* / N. Golyandina, V. Nekrutkin, A. Zhigljavsky. - Boca Raton: Chapman & Hall/CRC, 2001. - 305 p.
2. Hassani, Hossein. *Singular Spectrum Analysis: A Relatively New and Powerful Technique for Time series Analysis and Forecasting [Text]* / Hossein Hassani // *The 31st Annual Int. Sym. on Forecasting: The University of economics.* - Prague - 2011. - P. 1-11.
3. Данилов, Д.Л. *Главные компоненты временных рядов: метод «Гусеница» [Текст]* / Д.Л. Дани-

лов, А.А. Жигляевский. - СПб.: СпбГУ, 1997. - 150 с.

4. Айвазян, С.А. *Прикладная статистика и основы эконометрики [Текст]* / С.А. Айвазян, В.С. Мхитарян. - М.: ЮНИТИ, 1998. - 1006 с.

5. Орлов, Ю.Н. *Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков [Текст]* / Ю.Н. Орлов, К.П. Осминин. - М.: Editorial URSS, Книжный дом «ЛИБРОКОМ», 2011. - 384 с.

6. Афанасьев, В.Н. *Анализ временных рядов и прогнозирование [Текст]* / В.Н. Афанасьев, М.М. Юзбаев. - М.: Финансы и статистика, 2001. - 231 с.

7. *Применение метода локальной аппроксимации для прогноза экономических показателей [Текст]* / А.Ю. Лоскутов, Д.И. Журавлев, О.Л. Котляров // *Вопросы анализа и управления риском*, - 2003. - Т.1, № 1. - С. 21-31.

Поступила в редакцию 15.11.2011

Рецензент: д-р техн. наук, проф., зав. каф. «Системотехника» Э.Г. Петров, Харьковский национальный университет радиоэлектроники, Харьков.

ІДЕНТИФІКАЦІЯ СТРУКТУРИ НЕСТАЦІОНАРНОГО ЧАСОВОГО РЯДУ ЗА ДОПОМОГОЮ МЕТОДУ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛІЗУ

А.О. Чистякова, Б.В. Шамша

У роботі розглянутий метод сингулярного спектрального аналізу (SSA) для ідентифікації структури нестационарних часових рядів. Метою методу є виділення компонент часового ряду, таких як тренд та періодична складова. Рішення поставленої задачі необхідне для побудови моделі часового ряду і визначення неявних залежностей. Проведений аналіз структури нестационарного часового ряду ціни на цукор. Дані рекомендації з вибору параметрів при використанні методу SSA для рядів, які не можуть бути приведені до однорідних. Побудована модель нестационарного часового ряду з урахуванням компонент тренду та періодики.

Ключові слова: нестационарні часові ряди, ідентифікація моделі, сингулярний спектральний аналіз, головні компоненти.

IDENTIFICATION OF THE STRUCTURE OF NONSTATIONARY TIME SERIES WITH THE SINGULAR SPECTRUM ANALYSIS METHOD

A.A. Chistyakova, B.V. Shamsha

This paper represents the method of Singular Spectrum Analysis (SSA) to identify the structure of nonstationary time series. The purpose of this method is the selection of the time series components, such as trend and periodic component. Solution of this problem is necessary for constructing the model of time series and determination of the masked dependences. The analysis of the structure of nonstationary time series of prices of the sugar is carried out. Recommendations on the choice of parameters of SSA to identify the components of time series, which cannot be reduced to a uniform are given. The model of nonstationary time series taking into account the components of trend and periodicals is built.

Key words: nonstationary time series, identification of the model, singular spectrum analysis, principal components.

Чистякова Анна Александровна – аспирант каф. ИУС, Харьковский национальный университет радиоэлектроники, Харьков, Украина, e-mail: anewlove@mail.ru.

Шамша Борис Владимирович – канд. техн. наук, проф. каф. ИУС, Харьковский национальный университет радиоэлектроники, Харьков, Украина.