

ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ ОПРЕДЕЛЕНИЯ СОСТАВЛЯЮЩИХ НЕРЕГУЛЯРНЫХ ВРЕМЕННЫХ РЯДОВ ПРИ ПОМОЩИ СИНГУЛЯРНОГО РАЗЛОЖЕНИЯ

Проводиться розкладання тимчасового ряду на складові за допомогою сингулярного розкладання матриці розгортки. Аналізується ризик при побудові моделі прогнозу на підставі складових тимчасового ряду, використовуючи волатильність випадкової складової як чинник ризику

Ключові слова: часовий ряд, сингулярне розкладання, інформаційна технологія, математична модель

Проводится разложение временного ряда на составляющие при помощи сингулярного разложения матрицы развертки. Анализируется риск при построении модели прогноза на основании составляющих временного ряда, используя волатильность случайной составляющей как фактор риска

Ключевые слова: временной ряд, сингулярное разложение, информационная технология, математическая модель

Decomposition of temporal row is conducted on constituents through singular decomposition of matrix of involute. A risk at the construction of model of prognosis is analysed on the basis of constituents of temporal row using volatil'nost' of casual constituent as a risk factor

Keywords: temporal row, singular decomposition, information technology, mathematical model

А.А. Чистякова
Аспирантка*

Контактный тел.: 097-209-15-58
E-mail: anewlove@mail.ru

Д.С. Негурица*

Контактный тел.: 050-262-12-49
E-mail: neguritsa@gmail.com

Б.В. Шамша

Кандидат технических наук, профессор*
Контактный тел.: (057) 70-21-451
E-mail: iyc@kture.kharkov.ua

*Кафедра информационных управляющих систем
Харьковский национальный университет
радиоэлектроники
пр. Ленина 14, Харьков, Украина, 61166

1. Введение

Существует несколько подходов к оценке структуры временного ряда, которые имеют ограниченное применение, связанной со спецификой объекта прогнозирования. В работе рассматриваются нерегулярные временные ряды на примере курсов валют, которые характеризуются нелинейностью, нестационарностью, гистерезисом, хаотичностью и т.д. Структура модели определяется при помощи сингулярного разложения с оценкой случайной составляющей. При помощи ARCH модели оценивается прогноз волатильности случайной составляющей, которая в дальнейшем используется для определения риска при принятии решений по управлению на основе модели прогноза.

2. Постановка проблемы в общем виде

Заметим, что нерегулярность временных рядов заключается в наличии резких переходов, гистерезисом, изменчивостью структуры, детерминированным хаосом, нелинейностью, нестационарностью. В таких условиях обнаружение тренда, гармонических, случайных составляющих представляет определенную трудность, связанную с их неявной выраженностью. Известно, что если характер временного ряда линеен и математическая модель ряда довольно проста и легко интерпретируема оценить составляющие не представляет трудностей. Если же структура временного ряда нелинейна, а математическая модель выше первой степени, оценить наличие компонент ряда достаточно сложно.

В таком случае применяются специально разработанные методы. Нелинейные временные ряды, несмотря на различие в структуре, поведении, описании, характеризуются аналогиями в их организации и функционировании. С этой точки зрения представляет интерес рассматривать анализ нерегулярных временных рядов с точки зрения нелинейной динамики.

Успех решения задачи прогнозирования нерегулярных временных рядов в значительной степени зависит от наличия различных составляющих во временном ряде.

Известно что:

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

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. Для их определения в работе используем метод сингулярного разложения.

3. Основные этапы информационной технологии оценки структуры временного ряда

Исходя из специфики функционирования нерегулярных временных рядов, определим основные этапы построения информационной технологии оценки структуры исходной последовательности данных. Под информационной технологией анализа временного ряда будем понимать следующую последовательность действий:

- 1) получение набора исходного временного ряда;
- 2) трансформация данных с предварительной обработкой;
- 3) проверка статистических гипотез о наличии стационарности, нелинейности, гетероскедастичности, фильтрация данных временного ряда;
- 4) выбор метода оценки структуры нерегулярного временного ряда;
- 5) выбор структуры модели прогнозирования;
- 6) выбор метода построения модели прогнозирования, выбор метода идентификации параметров модели прогнозирования;
- 7) выбор глубины прогноза и построение аналитической модели прогноза;
- 8) оценка адекватности модели;
- 9) определение стратегии модернизации модели.

Основная цель данной работы состоит в оценке структуры временного ряда, что позволит выделить отдельные составляющие, провести их анализ и описать их математически с целью получения эффективной модели прогноза. Сингулярный анализ используется в данной работе для разделения сложного процесса на множество элементарных компонент, по которым в дальнейшем строится общая модель прогноза.

4. Основные этапы информационной технологии разложения исходного временного ряда на составляющие

Выделение составляющих временного ряда при помощи сингулярного разложения достигается преобразованием исходного временного ряда в несколько рядов и получения множества отрезков. Для решения этой задачи предложено представить исходный временной ряд в виде векторов определенной длины L , взятых последовательно.

Исходя из внутренних свойств системы для исходного ряда $F = (f_0, \dots, f_{N-1})$, состоящего из целого числа значений N и характеризующего движение нестационарной системы, подбирается целое число L — длина выборки (окна), $L, L \in \mathbb{N}, 1 < L < N$ и составляется матрица развертки X , содержащая $K = N - L + 1$ векторов развертки, имеющих заданную размерность L . То есть матрица имеет вид:

$$X = \begin{pmatrix} f_0 & f_1 & \dots & f_{N-L} \\ f_1 & f_2 & \dots & f_{N-L+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & \dots & f_{N-1} \end{pmatrix}, \quad (2)$$

... где L — число элементов окна, N — число значений исходного ряда, \vdots — множество значений исходного ряда F по вертикали, \ddots — множество значений исходного ряда F по горизонтали, \ddots — множество значений исходного ряда F по горизонтали и вертикали.

Далее производится сингулярное разложение матрицы X размером $L \times K$, которым называется её представление в виде:

$$X = U \times W \times V^T, \quad (3)$$

где U — ортогональная матрица размером $L \times L$, V — ортогональная матрица размером $K \times K$, W — матрица размером $L \times K$, на главной диагонали которой находятся сингулярные неотрицательные числа, расположенные в порядке убывания, а все не диагональные элементы равны нулю.

С учетом особенностей матрицы W для получения матрицы X требуется не L столбцов матрицы U , а лишь первые $\min(L, K)$ столбцов, аналогично, лишь первые $\min(L, K)$ строк матрицы V^T влияют на результат произведения. Эти столбцы и строки являются левыми и правыми сингулярными векторами.

Таким образом, исходная матрица развертки разлагается на матрицу левых сингулярных векторов U размерности $L \times M$:

$$M = \min(L, K), \quad (4)$$

транспонированную матрицу правых сингулярных векторов V^T размерности $M \times K$ и матрицу сингулярных чисел W размерности $M \times M$.

На третьем этапе исходя из предположения, что каждое сингулярное число W_i в (3) характеризует определенную составляющую нерегулярного временного ряда, можно определить степень влияния каждого сингулярного числа на результирующий ряд и выделить составляющие процесса.

Так для того, чтобы определить влияние первого сингулярного числа, нужно найти обратную матрицу развертки для первой составляющей по формуле:

$$X_{iL \times K}^0 = U_{L \times M} \times W_{iM \times i} \times V_{iK}^T = \begin{bmatrix} u_{11} & \dots & u_{1M} \\ u_{21} & \dots & u_{2M} \\ \vdots & \ddots & \vdots \\ u_{L1} & \dots & u_{LM} \end{bmatrix} \times \begin{bmatrix} w_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \times [v_{11} \dots v_{1K}] \quad (5)$$

Влияние совокупности i сингулярных чисел определяется по соотношению:

$$X_{iL \times K} = U_{L \times M} \times W_{iM \times i} \times V_{iK}^T = \begin{bmatrix} u_{11} & \dots & u_{1M} \\ u_{21} & \dots & u_{2M} \\ \vdots & \ddots & \vdots \\ u_{L1} & \dots & u_{LM} \end{bmatrix} \times \begin{bmatrix} w_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_i \end{bmatrix} \times \begin{bmatrix} v_{11} & \dots & v_{1K} \\ \vdots & \ddots & \vdots \\ v_{i1} & \dots & v_{iK} \end{bmatrix} \quad (6)$$

Из выражения (5) влияние последнего i -го сингулярного числа определяется по формуле:

$$X_{iL \times K}^0 = X'_{L \times K} - X'_{i-1L \times K} = U_{L \times M} \times W_{iM \times i} \times V_{iK}^T - U_{L \times M} \times W_{(i-1)M \times (i-1)} \times V_{(i-1)K}^T \quad (7)$$

Из соотношения (2) для возвращения к исходному ряду необходимо провести операцию усреднения по второстепенным диагоналям полученной суммарной матрицы X .

Аналогичным образом, для формирования каждой простейшей составляющей, определяемой соответствующим сингулярным числом, необходимо провести операцию усреднения по второстепенным диагоналям обратной матрицы развертки X_i^0 и получить ряд F_i .

$$\left. \begin{aligned} A_{L \times K}^0 &= U_{L \times M} \times W_{iM \times i} \times V_{iK}^T \rightarrow F \\ A'_{i-1L \times K} &= U_{L \times M} \times W_{(i-1)M \times (i-1)} \times V_{(i-1)K}^T \rightarrow F_{i-1} \end{aligned} \right\} \rightarrow F_i = F_i - F_{i-1} \quad (8)$$

Таким образом, показано, что с помощью метода сингулярного разложения исходную выборку нерегулярного временного ряда можно разложить на произвольное число аддитивных компонент, количество которых определяется длиной траекторной матрицы X .

5. Исследование структуры временного ряда курса валют

В качестве примера разложения нерегулярного временного ряда на составляющие рассмотрим ряд курса валют EUR/USD, история которого содержит 2014 значений, зафиксированных через каждые 5 минут. Временной ряд представлен на рис. 1.

Проведем разложение временного ряда курса валют, используя для этого формулы (1-8). Собственные числа после разложения упорядочиваются по убыванию, и анализируются самые большие, влияющие на собственные числа и вектора сингулярного разложения.



Рис. 1. Временной ряд курсов валют

Исследуя временной ряд вышеизложенным методом, возможно выделить такие компоненты как тренд и периодическая составляющая на основании выбранных значений сингулярного разложения. Для выделения тренда используются графики полученных упорядоченных собственных троек. Первые 9 из них представлены на рис. 2.

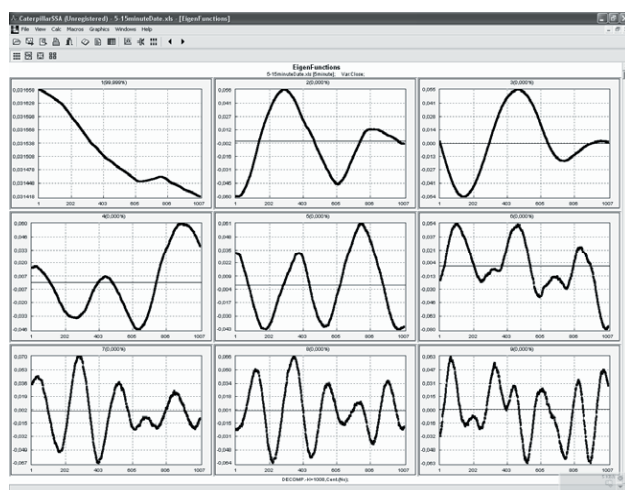


Рис. 2. Графики собственных троек сингулярного разложения

Так как тренду соответствует плавно меняющаяся составляющая, то для восстановления данной компоненты предлагается использовать собственные тройки под номерами 1, 2, 3.

Приступим к анализу и выделению периодической составляющей. Для этого рассмотрим двумерные диаграммы собственных троек, первые 9 из которых представлены на рис. 3. Пары собственных векторов, соответствующих модулированным гармоническим составляющим ряда, удобнее искать на двумерных диаграммах, на которых по одной оси откладываются элементы первого вектора из пары и по другой – элементы второго вектора. Из-за свойств собственных чисел, соответствующих гармоникам, нам достаточно рассматривать 2D диаграммы только соседних собственных векторов.

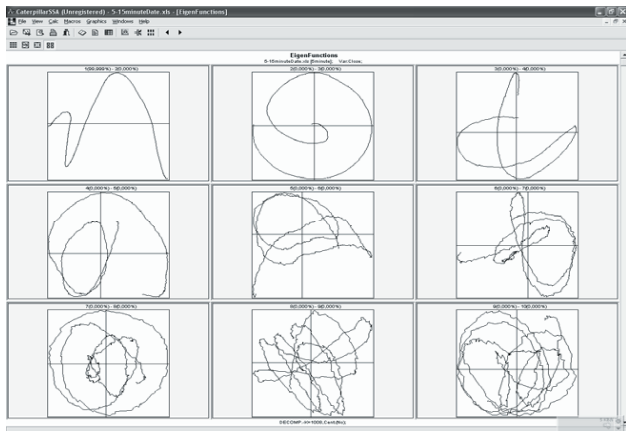


Рис. 3. Двумерные диаграммы собственных троек сингулярного разложения

Рассмотрев самые значимые из графиков было отмечено, что компонентами, которые соответствуют периодической составляющей являются номера 2, 3, 4, 5, 7, 8, 9, 10.

После восстановления поданным компонентам сингулярного разложения получим следующую модель, представленную на рис .4.

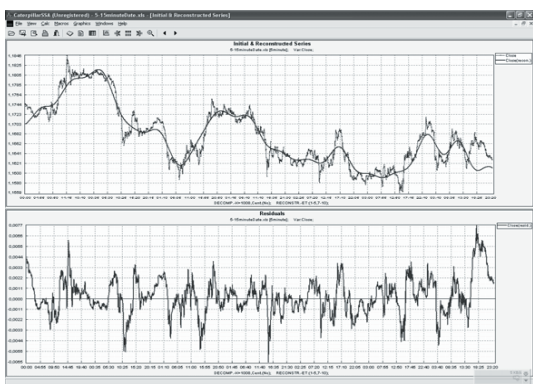


Рис. 4. Восстановленный временной ряд по выбранным компонентам разложения

На основе проведенных экспериментов определено, что качество модели, а также выделяемых компонент временного ряда зависит от длины окна L , которая определяет пространство исследования векторов временного ряда.

На сегодняшний день предложены некоторые критерии для оценки выделяемых трендовых и периодических компонент:

- критерий уровня гладкости H :

$$H = \max(\beta_i) : i = \overline{1, k-1}, \tag{9}$$

где β - угол приращения, характеризующий изменение значения производной в i -ой точке нестационарной выборки временного ряда, содержащего p значений;

- критерий уровня отклонения колебаний C :

$$C = \max(\Delta A_i) : i = \overline{1, m-1}, \tag{10}$$

где ΔA - приращение между соседними колебаниями (разница амплитуд см. графики 3), m – количество колебаний.

Данные критерии позволяют контролировать качество выделяемых сингулярных составляющих.

Следующим этапом построения модели прогнозирования является выбор метода прогнозирования всех полученных составляющих временного ряда. В настоящее время существует множество методов прогнозирования, каждый из которых основывается на своих предпосылках и предположениях. Известно, что тренд лучше всего прогнозировать методом скользящего среднего или экспоненциального сглаживания. В свою очередь, для построения модели периодической составляющей метод разложения Фурье дает хорошие результаты. Что касается случайной составляющей, то предлагается применить Generalized Autoregressive Conditional Heteroskedasticity (GARCH) модель, которая применяется для прогнозирования волатильности. Случайная составляющая временного ряда показывает волатильность, которая, в сущности, и является мерой неопределенности, риска. Таким образом, построив модель волатильности, можно оценить риск по принятию общей, суммарной модели прогноза.

Смысл GARCH-модели состоит в том, что если абсолютная величина случайной составляющей оказывается большой, то это приводит к повышению дисперсии в последующие периоды. При этом наблюдается и обратный эффект: повышение дисперсии означает увеличение вероятности появления больших ошибок модели. Данное свойство модели хорошо согласуется с реальными финансовыми временными рядами: на рынке наблюдаются как относительно стабильные периоды, так и весьма бурные. Этот эффект также называют кластеризацией волатильности.

Как пишет Мандельброт (Mandelbrot(1963)),
 "... большие изменения как правило сопровождаются большими изменениями, независимо от их знака, а малые изменения как правило сопровождаются малыми изменениями..."

Модель GARCH(1,1) задается следующими соотношениями

$$y_k = \sigma_k \epsilon_k, \quad \sigma_k^2 = \alpha + \beta y_{k-1}^2 + \gamma \sigma_{k-1}^2, \quad k \in Z, \tag{11}$$

где $\{\epsilon_k\}$ – независимые одинаково распределенные случайные величины с функцией распределения $G(x)$, $E\epsilon_k^2 = 1$; $a = (\alpha, \beta, \gamma)$ - вектор неизвестных параметров, $\alpha, \beta, \gamma > 0$, $\beta + \gamma < 1$. Тогда существует единственное строго стационарное и эргодическое решение $\{y_k\}$.

Одна из основных задач статистического анализа модели (11) – оценивание неизвестных параметров. Если функция $G(x)$ известна, несложно построить оценки максимального правдоподобия (ОМП) или им асимптотически эквивалентные.

Итак, для построения данной модели для случайной составляющей исходного временного ряда следует вычислить ее параметры. Для этого используются специальные программные средства, где уже имеются такие программные средства, как Matrixer. Модель GARCH с вычисленными коэффициентами приведена ниже:

$$\sigma_k^2 = 0.443 + 0.29y_{k-1}^2 + 0.229\sigma_{k-1}^2 \tag{12}$$

Полученная модель является моделью волатильности, которая в свою очередь определяет степень риска при прогнозировании. Построенная модель является адекватной, так как ошибка модели не превышает

10%. Т-статистика говорит о том, что все вычисленные коэффициенты являются значимыми. В соответствии с моделью случайной составляющей временного ряда и свойством кластеризации волатильности можно говорить об оценке риска при прогнозировании.

6. Выводы

В работе показано, что построение модели прогнозирования является многоэтапной процедурой, для проведения которой необходимо решение задач сбора данных и оценки их количества, предварительной статистической обработки, оценки статистических характеристик данных, проверки статистических гипотез, определения структуры временного ряда, выбора эффективного метода прогнозирования, проведения идентификации модели прогнозирования и оценка ее адекватности. Представлена информационная технология построения модели прогнозирования для нерегулярных временных рядов. В работе впервые пред-

ложена многоэтапная процедура построения модели прогнозирования с использованием оценки риска для интерпретации адекватности.

Литература

1. Engle R. Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation / *Econometrica*. 1982. V. 50. Pp. 987 – 1008.
2. Голяндина Н.Э. Метод «Гусеница»-SSA: анализ временных рядов: учебн. пос. / Н.Э. Голяндина. – СПб: Изд-во СПбГУ, 2004. – 76 с.
3. Кедрин В.С. Моделирование нестационарных процессов на базе алгоритмов сингулярного разложения с использованием систем искусственного интеллекта. / В.С. Кедрин, М.К. Сальникова // Моделирование. Теория, методы и средства: Материалы VII Междунар. науч.-практ. конф., г. Новочеркасск, 6 апр. 2007 г.: В 3 ч. / Юж.-Рос. гос. техн. ун-т (НПИ). – Новочеркасск:ЮРГТУ, 2007. – Ч.1. – С. 37-43.

В статті розглядаються методичні питання проектування раціональних систем управління міськими пасажирськими перевезеннями

Ключові слова: управління системами міських пасажирських перевезень, пасажиропотік, пасажирська кореспонденція, транспортна мережа

В статье рассматриваются методические вопросы проектирования рациональных систем управления городскими пассажирскими перевозками

Ключевые слова: управление системами городских пассажирских перевозок, пассажиропоток, пассажирская корреспонденция, транспортная сеть

Methodological principles of projection rational management of urban passenger transportation are considered in the article

Key words: control of urban passenger transport, urban passenger transport, passenger, passenger correspondence

Актуальність теми

Розвиток міських пасажирських перевезень (далі – МПП) в умовах кардинальних економічних, соціальних, демографічних змін, які спостерігаються в останні десятиліття в Україні, з одного боку, і поширення інформаційних та телекомунікаційних технологій,

з іншого, вимагають постановки нових і переопрацювання ряду відомих наукових задач щодо раціонального управління процесами і системами МПП на стратегічному, тактичному та оперативному рівнях. Проведений аналіз показав, що найбільш актуальними щодо подальших досліджень є наступні питання: розроблення методів, моделей та алгоритмів побудови

УДК 656.13

КОНЦЕПЦІЯ ПРОЕКТУВАННЯ СИСТЕМ МІСЬКИХ ПАСАЖИРСЬКИХ ПЕРЕВЕЗЕНЬ

І. Ф. Шпильовий

Заступник начальника

Головне управління транспорту та зв'язку

Київська міська державна адміністрація

вул. Леонтовича, 6, м. Київ, Україна

Контактний тел.: 067-607-62-48