УДК 004.942: 519.866

**Д.С. РЕВЕНКО**

***Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»***

# МЕТОД КОМПЛЕКСНОЙ ИДЕНТИФИКАЦИИ И ПРОГНОЗИРОВАНИЯ СТРУКТУРНЫХ КОМПОНЕНТ ДИНАМИЧЕСКИХ ПРОЦЕССОВ С НЕОПРЕДЕЛЕННЫМИ ДАННЫМИ

*В статье рассматриваются динамические экономические процессы с неопределенными данными, когда информация об экономическом явлении (процессе) ограничена либо недоступна, в таких случаях знания о процессе могут быть заменены на диапазон информации (интервал). В качестве метода исследования таких динамических процессов предложен метод комплексной идентификации и прогнозирования структурных компонент динамического интервального ряда. Сущность данного метода заключается в представлении интервального динамического ряда в виде суммы компонент: – основной тенденции (тренда), сезонной и случайной компонент.*

***Ключевые слова:*** *динамические экономические процессы, неопределенность данных, интервалы, тренд, сезонная составляющая, экспоненциальное сглаживание.*

При изучении временных (динамических) рядов экономических данных, существует широкий класс задач, обладающих неоднозначностью и неопределенностью, в таких случаях полезно использовать аппарат интервального анализа, когда явление можно описать с помощью диапазона изменения (интервала) в каждый момент времени $t$.

В последнее время интервальному описанию факторов неопределенности уделяется все больше внимание, как наиболее адекватно отвечающему широкому классу экономических задач с неопределенными данными.

Типичными структурными компонентами динамического интервального ряда в краткосрочном и среднесрочном разрезе являются: основная тенденция (тренд) $T$, циклическая (сезонная) составляющая $[S]$ и случайная компонента $[e]$, следовательно, динамически интервальный ряд $[y]$ можно представить в виде суммы перечисленных компонент (рис. 1):

$$[y] = T + [S] + [e]. \tag{1}$$

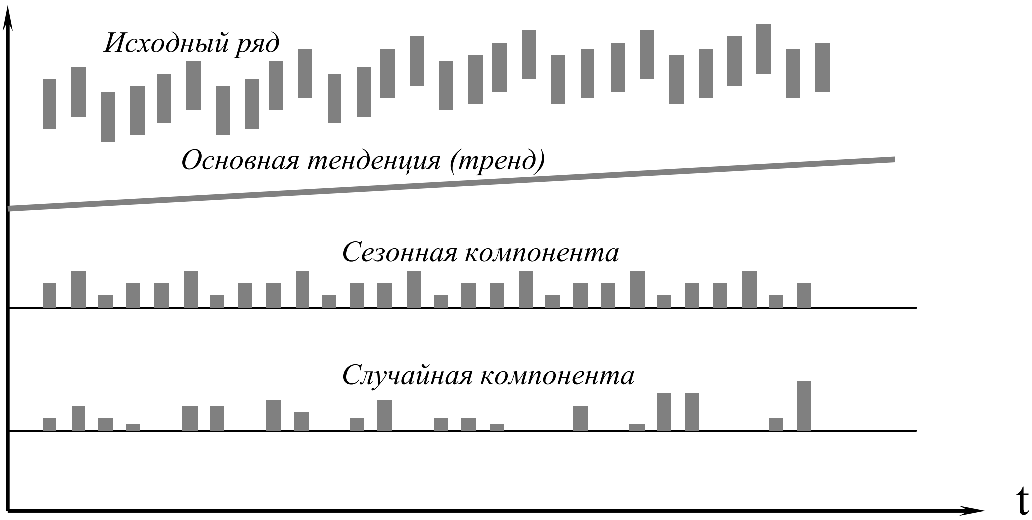


Рис. 1. Структура интервального динамического ряда.

Использование интервального подхода для решения задач описания факторов неопределенности обладает своей спецификой, что несет в себе совершенно новые подходы для решения задач данного класса. В нашем случае, для разработки метода идентификации и прогнозировании структурных компонент интервальных динамических рядов.

Специфика разработанного метода состоит в том, что интервальный динамических ряд представляют в виде суммы структурных компонент (тренда, сезонной  и случайной компоненты) и поэтапном их выделении из исходного ряда и в последующем прогнозировании каждой из компонент и восстановлении ряда.

На первом этапе нужно выяснить, существует ли основная тенденция в исследуемом процессе. На практике наиболее часто используется «критерий серий» [1] для проверки «наличия – отсутствия» тренда.

Критерий серий основан на медиане выборки, который может быть представлен в виде следующей последовательности шагов.

1) из исходного интервального ряда со средними уровнями $\mathrm{mid}[x_1], \mathrm{mid}[x_2], ..., \mathrm{mid}[x_i]$ образуется ранжированный вариационный ряд для $\mathrm{mid}[x_1]', \mathrm{mid}[x_2]', ..., \mathrm{mid}[x_i]'$, где $\mathrm{mid}[x_1]'$ – наименьшее значение из уровней исходного ряда $\mathrm{mid}[x_1], \mathrm{mid}[x_2], ..., \mathrm{mid}[x_i]$;

2) определяется медиана ( $Me_{mid}$ ) этого вариационного ряда. В случае нечетного значения длины $t$ $(t = 2m + 1)$ $Me_{mid} = \mathrm{mid}[x_{m+1}]'$, в противном случае ( $t = 2m$ ) $Me_{mid} = \left(\mathrm{mid}[x_m]' + \mathrm{mid}[x_{m+1}]'\right)/2$;

3) образуется последовательность $\delta_n$ из плюсов и минусов по следующему правилу:

$$\delta_n = \begin{cases} +, \text{если } \mathrm{mid}[x_i] > Me_{mid}; \\ -, \text{если } \mathrm{mid}[x_i] < Me_{mid}. \end{cases} \tag{2}$$

Если значение уровня исходного ряда $\mathrm{mid}[x_i]$ равно медиане, то это значение пропускается. Очевидно, что общее число знаков «+» и «-» заранее неизвестно. Индекс $n$ может принимать значение $1, 2, ..., k$, где $k \le n$;

4) подсчитывается число серий $v(n)$ в совокупности $\delta_n$, где под серией понимается последовательность подряд идущих плюсов или минусов. Один плюс или один минус тоже будет считаться серией;

5) определяется $\tau_{max}$ – протяженность самой длинной серии;

6) проверка гипотезы основывается на том, что при условии случайности ряда (при отсутствии систематической составляющей) протяженность самой длинной серии не должна быть слишком большой, а общее число серий – слишком маленьким. Поэтому для того чтобы не была отвергнута гипотеза о случайности исходного ряда (об отсутствии систематической составляющей), должны выполняться следующие неравенства:

$$\begin{gathered} v(n) > \left[\frac{1}{2}\left(n + 1 - 1,96\sqrt{n-1}\right)\right], \\ \tau_{max} < \left[1,43 \cdot \ln(n+1)\right] \end{gathered} \tag{3}$$

где $n$ – длина временного ряда; $v(n)$ – число серий; $\tau_{max}$ – число подряд идущих плюсов или минусов в самой длинной серии.

Если хотя бы одно из неравенств (3) нарушается, то подтверждается наличие зависящей от времени неслучайной составляющей.

Квадратные скобки в правой части неравенства означают целую часть числа. Напомним, что целая часть числа – это целое число, ближайшее к нему и не превосходящее его [1].

Следующим этапом выделяем линейный стенд. Выбор линейного тренда вызван «принципом простоты» [2], а также информация, получаемая в условиях неопределенности, не имеет большого числа серий.

Выделение линейного тренда из интервального ряда необходимо проводить по средним значениям интервалов временного ряда.

Уравнение линейного движения описывается уравнением тренда вида:

$$\tilde{x}_i = a + b \cdot t_k, \tag{4}$$

где $\tilde{x}_i$ – выровненные, т.е. лишенные колебаний, уровни тренда для $i$-го наблюдения; $a$ – свободный член уравнения, число равных среднему выровненному уровню для момента или периода времени, принято за начало отсчета, т.е. для $t = 0$; $b$ – средняя величина изменения уровней ряда за единицу изменения времени; $t_k$ – номер моментов или периодов времени, к которым относится уровни временного ряда (год, квартал, месяц, дата).

Для вычисления параметров тренда наиболее часто используется метод наименьших квадратов, на основании которого можно вычислить $a$ и $b$ [3].

$$a = \frac{\sum mid[x_i] \cdot \sum t^2 - \sum t \cdot \sum mid[x_i]}{n \cdot \sum t^2 - \left(\sum t\right)^2}; \tag{5}$$

$$b = \frac{n \cdot \sum mid[x_i] \cdot t - \sum t \cdot \sum mid[x_i]}{n \cdot \sum t^2 - \left(\sum t\right)^2}. \tag{6}$$

После чего строится теоретический ряд и выделяется из границ исходного динамического ряда.

В рядах динамики, уровни которых являются месячными или квартальными показателями, наряду со случайными колебаниями часто наблюдаются сезонные колебания, под которыми понимается периодически повторяющееся из года в год повышение и снижение уровней в отдельных месяцах, кварталах. Сезонным колебаниям подвержены внутригодовые уровни многих показателей.

При изучении рядов динамики, содержащих «сезонную волну» используется ряд методов для решения этой задачи. Все они основаны на сравнении фактических уровней каждого месяца (или квартала) со средним уровнем, предполагающим равномерное распределение годового показателя по месяцам (либо кварталам) [3]. Эти показатели называются индексами сезонности:

$$I_s = \frac{\mathrm{mid}\left[x_i\right]}{Av} \cdot 100\% . \tag{7}$$

Если имеется тренд – сезонный ряд, тогда среднюю интервального ряда необходимо заменить на трендовый теоретический ряд:

$$I_s = \frac{\mathrm{mid}\left[x_i\right]}{\tilde{x}_i} \cdot 100\% . \tag{8}$$

Этот метод довольно прост, но в силу элемента случайности месячные данные одного года недостаточно надежны для определения меры сезонных колебаний. Поэтому рекомендуется пользоваться месячными (или квартальными) данными за ряд лет (в основном 3 года, хотя не исключена возможность использования данных за 2 года). В этом случае составляется матрица сезонных компонент:

$$\begin{matrix} I_{s11} & I_{s12} & \cdots & I_{s1n} \\ I_{s21} & I_{s22} & \cdots & I_{s2n} \\ \cdots & \cdots & \cdots & \cdots \\ I_{sn1} & I_{sn2} & \cdots & I_{snm} \end{matrix} , \tag{9}$$

где $n$ – длина систематического периода (неделя, месяц, квартал, год) с сезонной составляющей; $m$ – количество равных разбиений исходного ряда.

Следовательно, среднее столбцов матрицы (9) и будет корректным индексом сезонности.

После исследования ряда на наличие сезонной составляющей и определения ее параметров необходимо ее выделить из ряда полученного после выделения тренда. В итоге получен ряд, не имеющий системных составляющих, т.е. стационарный интервальный ряд. Исследование и прогнозирование остаточного интервального динамического ряда можно провести на основе интервальной модели экспоненциального сглаживания (10), особенность использования этой модели подразумевает решение следующих задач:

– поиск начального условия сглаживания;

– поиск константы сглаживания $\left[\underline{\alpha},\overline{\alpha}\right]$.

$$\left[\underline{F_t},\overline{F_t}\right]=\left[\underline{\alpha},\overline{\alpha}\right]\cdot\left[\underline{A_{t-1}},\overline{A_{t-1}}\right]+\sum_{i=1}^{t-2}\left[\underline{\alpha},\overline{\alpha}\right]\cdot(1-\left[\underline{\alpha},\overline{\alpha}\right])^i\cdot\left[\underline{A_{t-(i+1)}},\overline{A_{t-(i+1)}}\right], \quad (10)$$

где $\left[\underline{F_t},\overline{F_t}\right]$ – интервальное прогнозное значение показателя; $\left[\underline{A_{t-1}},\overline{A_{t-1}}\right]$ – текущий интервальный показатель прошлого периода; $\left[\underline{\alpha},\overline{\alpha}\right]$ – константа сглаживания ($0<\alpha<1$), представлена как врожденный интервал; t – количество рассматриваемых периодов.

При малых рядах предлагается использование следующего соотношения для определения начального условия $A_0$ для нижней и верхней границ интервалов соответственно:

$$\underline{A_0}=\frac{\underline{A_1}+\underline{A_2}+\underline{A_3}}{3};\ \overline{A_0}=\frac{\overline{A_1}+\overline{A_2}+\overline{A_3}}{3}. \quad (11)$$

Для определения параметра сглаживания $[\alpha]$, разработан эффективный метод, основанный на ретроспективном анализе исходных данных путем решения обратной задачи прогнозирования (соотношения 12 – 14): задание глубина ретроспективного анализа, определение константы сглаживания для прогноза на предыдущий период (t-1) путем решения полиномиального уравнения относительно $[\alpha_{t-1}]$:

$$\begin{gathered}\left[\underline{A_{t-1}},\overline{A_{t-1}}\right]=\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right]\cdot\left[\underline{A_{t-2}},\overline{A_{t-2}}\right]+\\+\sum_{i=1}^{t-2}\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right]\cdot(1-\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right])^i\cdot\left[\underline{A_{t-(i+1)}},\overline{A_{t-(i+2)}}\right],\end{gathered} \quad (12)$$

где $\left[\underline{A_{t-1}},\overline{A_{t-1}}\right]$ – реальное значение параметра в момент времени $t-1$, n – длительность временного ряда. Если уравнение имеет вещественные корни на интервале $\alpha\in[0,1]$, тогда это значит, что найденное значение $[\alpha_{t-1}]$ позволило бы получить совершенно точный прогноз на момент времени $t-1$. В качестве метода поиска корней полиномиального уравнения заданного интервальными коэффициентами можно использовать интервальный метод Ньютона [4]. После чего проводится последовательное повторение первого этапа по всей или части временной серии, получаем ряд полиномиальных уравнений с интервальными коэффициентами:

$$\left[\underline{A_{t\text{-}n+1}},\overline{A_{t\text{-}n+1}}\right]=\left[\underline{\alpha_{t-n+1}},\overline{\alpha_{t-n+1}}\right]\cdot\left[\underline{A_{t-n}},\overline{A_{t-n}}\right];$$

$$\ldots\ldots\ldots,$$

$$\left[\underline{A_{t\text{-}2}},\overline{A_{t\text{-}2}}\right]=\left[\underline{\alpha_{t-2}},\overline{\alpha_{t-2}}\right]\cdot\left[\underline{A_{t-3}},\overline{A_{t-3}}\right]+$$
$$+\sum_{i=1}^{t-3}\left[\underline{\alpha_{t-2}},\overline{\alpha_{t-2}}\right]\cdot(1-\left[\underline{\alpha_{t-2}},\overline{\alpha_{t-2}}\right])^{i}\cdot\left[\underline{A_{t-(i+3)}},\overline{A_{t-(i+3)}}\right]; \tag{13}$$
$$\left[\underline{A_{t\text{-}1}},\overline{A_{t\text{-}1}}\right]=\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right]\cdot\left[\underline{A_{t-2}},\overline{A_{t-2}}\right]+$$
$$+\sum_{i=1}^{t-2}\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right]\cdot(1-\left[\underline{\alpha_{t-1}},\overline{\alpha_{t-1}}\right])^{i}\cdot\left[\underline{A_{t-(i+1)}},\overline{A_{t-(i+2)}}\right].$$

Таким образом, имеем последовательность значений $\alpha$ для всех моментов временной серии, кроме первого:

$$\{\alpha\}=\{\alpha_{t\text{-}1},\alpha_{t\text{-}2},\alpha_{t\text{-}3},...,\alpha_{t\text{-}n+1},\} \tag{14}$$

В конце, интерполируем полученную тенденцию (14) на момент времени $t$ для получения прогнозного значения $\alpha_t$ [5], после чего осуществляем прогноз на шаг вперед и определяем прогнозный интервал $[\tilde{x}]$. Восстанавливаем динамический интервальный ряд, поэтапно добавляя прогнозные значения структурных компонент.

Оцениваем качество прогнозной модели. Качество прогноза характеризуют такие широко известные в литературе термины, как точность и надежность. Хотя содержание этих терминов часто раскрывают достаточно неоднозначно. Это можно объяснить тем, что пока не найден эффективный подход к оцениванию качества прогноза, кроме его практического подтверждения. О точности прогноза принято судить по размеру ошибки прогноза – разница между прогнозными и фактическими значениями исследуемого показателя. Показатель, оценивающий среднее расхождение между фактическими и прогнозными интервальными состояниями исследуемого показателя:

$$ML=\sqrt{\frac{\sum\limits_{t=n\text{-}m+1}^{n}\left(\underline{A_t}-\underline{F_t^T}\right)^2+\left(\overline{A_t}-\overline{F_t^T}\right)^2}{m}}, \tag{16}$$

где $m$ – количество шагов прогнозирования (аппроксимации); $\underline{A_t}$ и $\overline{A_t}$ – фактическое состояние нижней и верхней границ прогнозируемого интервального показателя; $\underline{F_t^T}$ и $\overline{F_t^T}$ – теоретическое (прогнозируемое) состояние верхней и нижней границ интервального показателя.

Относительный показатель (в процентах) среднего расхождения фактических данных и прогнозируемых к ширине исследуемого интервала:

$$RML = \sqrt{\frac{100}{m}\sum_{t=n-m+1}^{n}\frac{\left(\underline{A_t}-\underline{F_t^T}\right)^2+\left(\overline{A_t}-\overline{F_t^T}\right)^2}{\left(\overline{A_t}-\underline{A_t}\right)}}\,. \qquad (17)$$

Чем меньше значение описанных выше показателей, тем выше качество ретропрогноза.

Предложенный метод реализован инструментальными средствами интегрированного математического пакета MAPLE (библиотека для интервальных вычислений IntpaX), обеспечивающего все аналитические вычисления и необходимые средства визуализации. Апробация метода велась на основе данных о потерях электроэнергии в Харьковском регионе, оценка качества прогнозной модели показала результат в 10,94%, что является хорошим показателем по шкале качества прогноза.

## Литература

1. *Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва. – М.: ЮНИТИ-ДАНА, 2003. – 206 с.*

2. *84. Бокс Дж. Анализ временных рядов. Прогноз и управление. / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1972. – 194 с.*

3. *Громыко Л.Г. Теория статистики / Г.Л. Громыко. – М.: ИНФРА-М, 2002. – 414 с*

4. *Добронец Б.С. Интервальная математика: учебн. пособие / Б.С. Добронец. – Красноярск: Краснояр. гос. ун-т, 2004. – 216 с.*

5. *Вартанян В.М. Інформатизація економіки і управління: комп'ютерні інформаційні системи: навч. посібник / В.М. Вартанян, Л.О. Філіпковська. – Х.: Нац. аерокосм. ун-т «Харк. авіа. ін-т», 2008. – 184 с.*

**Рецензент:** д-р экон. наук, проф., директор **А.Н. Кизим**, Научно-исследовательский центр индустриальных проблем развития НАН Украины, Харьков.

## МЕТОД КОМПЛЕКСНОЇ ІДЕНТИФІКАЦІЇ ТА ПРОГНОЗУВАННЯ СТРУКТУРНИХ КОМПОНЕНТ ДИНАМІЧНИХ ПРОЦЕСІВ З НЕВИЗНАЧЕНИМ ДАНИМИ

*Д.С. Ревенко*

У статті розглянуті динамічні економічні процеси з невизначеними даними, коли інформація про економічне явище (процес) обмежена або недоступна, в таких випадках знання про процес можуть бути замінені на діапазон інформації (інтервал). В якості методу дослідження таких динамічних процесів запропонований метод комплексної ідентифікації та прогнозування структурних компонент динамічного інтервального ряду. Сутність даного методу полягає в представленні інтервального динамічного ряду у вигляді суми компонент: основної тенденції (тренду), сезонної і випадкової компонент.

**Ключові слова:** динамічні економічні процеси, невизначенні дані, інтервали, тренд, сезонна складова, експоненційне згладжування.

## METHOD FOR COMPREHENSIVE IDENTIFICATION AND PREDICTION OF STRUCTURAL COMPONENTS OF DYNAMIC PROCESSES WITH UNCERTAIN DATA

***D.S. Revenko***

The article deals with the dynamic economic processes with uncertain data, where information on an economic phenomenon (process) is limited or unavailable; in such cases, knowledge about the phenomenon can be replaced by intervals of information. We propose a comprehensive method of identifying and predicting structural components of the dynamic interval series as a method of investigating these dynamic processes. This method is based on representing a dynamic interval series as a sum of the main trend, seasonal component and random component.

**Keywords:** dynamic economic processes, data uncertainty, intervals, trend, seasonal component, exponential smoothing.

**Ревенко Даниил Сергеевич** – аспират кафедры экономики и маркетинга Национального аэрокосмического университета им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: revenko_dan@ukr.net.