

Метод главных компонент для взвешенных данных в процедуре многомерного статистического прогнозирования*

Цель исследования. Предположим, что исследуется динамика состояния некоторого объекта. Его состояние описывается системой заданных показателей. Среди них могут быть показатели, являющиеся линейной комбинацией других. Целью любой процедуры прогнозирования является решение двух задач: во-первых, оценка ожидаемого прогнозного значения, во-вторых, оценка доверительного интервала для возможных других прогнозных значений. Процедура прогнозирования многомерная. Так как показатели описывают один и тот же объект, то кроме явных зависимостей среди них могут быть и не явные. Метод главных компонент эффективно учитывает вариацию данных в системе исследуемых показателей. Следовательно, данный метод желательно использовать в процедуре прогнозирования. Итоги прогнозирования были бы более адекватными, если бы была возможность реализовывать разные стратегии прогнозирования. Но это потребует модификации традиционного метода главных компонент. Поэтому это является главной целью данного исследования. Сопутствующей целью является исследование возможности решения второй задачи прогнозирования, которая является сложнее первой.

Материалы и методы. При оценке доверительного интервала необходимо обозначить процедуру оценки ожидаемого прогнозного значения. При этом полезно было бы использовать методы многомерных временных рядов. Обычно при этом различные модели временного ряда используют понятие временного лага. Их количество и весовая значимость в модели может быть разной. В данном исследовании предлагается модель временного ряда на основе метода экспоненциального сглаживания. Процедура прогнозирования многомерная. Она будет опираться на правило согласованного изменения данных. Поэтому алгоритм прогнозной оценки отдельного показателя представлен в виде,

который будет удобен для построения и практической использования данного правила в будущем. Метод главных компонент должен учитывать веса значений показателей. Это необходимо для реализации различных стратегий оценки границ интервала прогнозных значений. Предлагаемая стандартизация взвешенных данных способствует выполнению основной теоремы факторного анализа. За счет этого обеспечивается построение ортонормированного базиса в факторном пространстве. При этом не потребовалось строить итерационный алгоритм, характерный для подобных исследований.

Результаты. Для тестового набора данных проведены сравнительные расчеты традиционным и взвешенным методом главных компонент. Он показывает, что основные характеристики компонентного анализа сохраняются. Один из рассматриваемых показателей явно зависит от других. Поэтому оба метода показывают, что количество факторов меньше чем количество показателей. Все показатели имеют хорошую связь с факторами. В традиционном методе зависимый показатель входит в первую главную компоненту. В модифицированном методе этот показатель лучше связан со второй компонентой.

Заключение. Было показано, что элементы факторной матрицы, соответствующие прогнозируемому времени, могут быть выражены как средневзвешенные предыдущих факторных значений. Это позволит оценить границы доверительного интервала для каждого отдельного показателя, а также для комплексного показателя всей системы. При этом учитывается как согласованность изменения данных, так и стратегия прогнозирования.

Ключевые слова: взвешенный метод главных компонент, многомерное статистическое прогнозирование

Viktor V. Nikitin, Dmitriy V. Bobin

Chuvash State University named after I.N. Ulyanov, Cheboksary, Russia

Principal Component Analysis for Weighted Data in the Procedure of Multidimensional Statistical Forecasting

Purpose of the research. Let's assume that the dynamics of the state of some object is being investigated. Its state is described by a system of specified indicators. Among them, some may be a linear combination of other indicators. The aim of any forecasting procedure is to solve two problems: first, to estimate the expected forecast value, and second, to estimate the confidence interval for possible other forecast values. The prediction procedure is multidimensional. Since the indicators describe the same object, in addition to explicit dependencies, there may be hidden dependencies among them. The principal component analysis effectively takes into account the variation of data in the system of the studied indicators. Therefore, it is desirable to use this method in the forecasting procedure. The results of forecasting would be more adequate if it were possible to implement different forecasting strategies. But this will require a modification of the traditional principal component analysis. Therefore, this is the main aim of this study. A related

aim is to investigate the possibility of solving the second forecasting problem, which is more complex than the first one.

Materials and research methods. When estimating the confidence interval, it is necessary to specify the procedure for estimating the expected forecast value. At the same time, it would be useful to use the methods of multidimensional time series. Usually, different time series models use the concept of time lag. Their number and weight significance in the model may be different. In this study, we propose a time series model based on the exponential smoothing method. The prediction procedure is multidimensional. It will rely on the rule of agreed upon data change. Therefore, the algorithm for predictive evaluation of a particular indicator is presented in a form that will be convenient for building and practical use of this rule in the future. The principal component analysis should take into account the weights of the indicator values. This is necessary for the implementation of various strategies for estimating the boundaries

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №19-410-210007р_а

of the forecast values interval. The proposed standardization of weighted data promotes to the implementation of the main theorem of factor analysis. This ensures the construction of an orthonormal basis in the factor area. At the same time, it was not necessary to build an iterative algorithm, which is typical for such studies.

Results. For the test data set, comparative calculations were performed using the traditional and weighted principal component analysis. It shows that the main characteristics of the component analysis are preserved. One of the indicators under consideration clearly depends on the others. Therefore, both methods show that the number of factors is less than the number of indicators. All indicators have a good relationship with the factors. In the traditional method, the dependent

indicator is included in the first main component. In the modified method, this indicator is better related to the second component.

Conclusion. It was shown that the elements of the factor matrix corresponding to the forecast time can be expressed as weighted averages of the previous factor values. This will allow us to estimate the limits of the confidence interval for each individual indicator, as well as for the complex indicator of the entire system. This takes into account both the consistency of data changes and the forecasting strategy.

Keywords: weighted principal component analysis, multidimensional statistical forecasting.

Введение

К основным задачам прогнозирования можно было бы отнести следующие: во-первых, оценка ожидаемого прогнозного значения интересующих нас показателей, во-вторых, оценка доверительного интервала для других возможных значений данных показателей. Данная работа имеет своей целью способствовать решению второй задачи на основе привлечения метода главных компонент. Однако при этом необходимо указать подходы по решению первой.

Подходы по определению прогнозных значений можно было бы свести к двум в зависимости от характера проводимых исследований. Если исследуются однотипные объекты по заданному набору показателей, то не редко применяются линейные модели многомерной регрессии [1, 2]. Модель выражает определенную точку зрения исследователя на анализируемый объект. Поэтому они могут несколько отличаться. Используя метод главных компонент можно их объединить в одну более общую [3]. Если исследуется динамика состояния объекта, то анализ опирается на понятие о многомерных временных рядах. Базовые положения подобного анализа хорошо представлены в монографии Э. Хеннана [4]. Часто в моделях временных рядов используются временные лаги. Сколько их входит в модель, с какими коэффициентами, каково правило согласованной динамики определяется характером конкретного исследования [5, 6, 7].

В данной работе предполагается, что исследуется динамика состояния некоторого объекта. Прогнозные значения по временным рядам предлагается оценивать на основе метода экспоненциального сглаживания. Это способствует формулировке удобного к практической реализации принципа согласованного изменения данных. Однако этот вопрос в данной работе не обсуждается.

Экспоненциальное сглаживание в сочетании с использованием метода главных компонент способствует поиску решения второй задачи прогнозирования. Как известно, данный метод опирается на корреляционную матрицу данных, в которой находят свое отражение, как их вариация, так и их взаимное влияние. Поэтому эти

особенности автоматически отразятся на искомым доверительных интервалах. Желательно учесть различные сценарии прогнозирования, а это потребует модификации традиционного метода главных компонент.

Общеизвестно, что метод главных компонент берет начало с работы К. Пирсона [8], в которой он обосновал возможность эффективного анализа изменчивости исходных данных за счет перехода в новое факторное пространство. Однако, данные могут быть не слишком аккуратно сгруппированы, как это было представлено в его работе. Поэтому появились многочисленные модификации данного метода, хороший обзор которых представлен в монографии [9], подготовленный международным коллективом авторов. Одной из первых появилась модификация, связанная с тем, что данные располагаются не вдоль оси воображаемого эллипсоида, а явно группируются вдоль некоторой кривой. Это и послужило появлению нелинейных методов главных компонент [10, 11]. Кроме того, данные могут представлять собой совокупность изолированных множеств, что привело к появлению метода независимых главных компонент [12, 13, 14]. Данные могут составлять не изолированные множества, но при этом будет наблюдаться несколько сгущений. В этом случае метод главных компонент модифицировался в кластерный анализ [15, 16].

Еще одна из модификаций данного метода позволила решать задачи, связанные с анализом нейронных сетей [17]. При этом возникает необходимость использовать веса, но не самих элементов данных, а матриц. Непосредственно веса данных в ходе применения метода главных компонент рассматриваются в работах [18, 19]. При этом они не задаются, а итерационно вычисляются в ходе робастного анализа данных. Элементы с наименьшими весами будут считаться «плохими» и подлежать удалению.

Ближе всего к тематике данной статьи относятся работы [20, 21, 22]. В работе [20] веса лишь косвенно связаны со значениями показателей. Решается задача о восстановлении первоначальных значений показателей. Изменяющиеся веса присутствуют в итерационном процессе корректировки матрицы факторного отображения.

При этом целью ставится снижение отличий между первоначальными и восстановленными значениями показателей согласно некоторой метрике. В следующей работе исследуется ситуация, когда матрица исходных данных может иметь в себе заведомые неточности (измерительный шум). Это приводит к неверным оценкам ковариационной матрицы, лежащей в основе метода главных компонент. Веса вводятся для элементов ковариационной матрицы, по отношению к которой итерационным способом отыскивается ортонормированный базис факторного пространства. В работе [22] веса являются характеристиками значений заданной матрицы показателей. Веса, например, могут появиться при свертке большой базы данных по принципу – как часто встречается. При этом по некоторым данным веса могут равняться нулю. Поиск ортонормированного базиса здесь, как и в предыдущей работе, есть результат реализации некоторого итерационного процесса. Эти работы не касались вопросов прогнозирования динамических процессов. Поэтому матрица данных оставалась неизменной.

В данном исследовании рассматривается единая стратегия прогнозирования по всей системе показателей. Поэтому веса данных одинаковые по всем показателям. Предлагается провести специфическую стандартизацию взвешенных данных. Она способствует выполнению основной теоремы факторного анализа и позволяет рассчитать ортонормированный базис факторного пространства, не прибегая к построению итерационной процедуры. Проводится сравнение традиционного и предлагаемого модифицированного метода главных компонент. Показано, что элементы факторной матрицы на прогнозный момент времени определяются средневзвешенной предшествующих значений данной матрицы. Это позволит оценить доверительные интервалы как для каждого отдельного показателя, так и для интегрального показателя всей системы, основываясь на их главных компонентах.

Начальная фаза прогнозирования

Допустим, что динамика состояния исследуемого объекта характеризуется m показателями $X_j, j=1, m$, по которым имеются статистические данные за n последовательных моментов времени. Необходимо спрогнозировать значение этих показателей в следующий $(n+1)$ момент времени. Представим имеющиеся и прогнозные значения $X_j, j=1, m$ в виде матрицы

$$\mathbf{X}_{(n+1) \times m} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \\ x_{(n+1)1}^{\text{пр}} & \dots & x_{(n+1)m}^{\text{пр}} \end{pmatrix}. \quad (1)$$

Допустим, прогнозные значения по каждому показателю оцениваются следующим способом

$$x_{(n+1)j}^{\text{пр}} = x_{(n+1)j}^{\text{ср}} + \Delta x_{(n+1)j}, \quad j=\overline{1, m}, \quad (2)$$

где первое слагаемое есть среднее значение показателя за предыдущие моменты времени, а второе трендовое приращение от этого среднего. При этом, оценивание прогнозного значения подобно использованию одномерной линейной регрессии.

Для дальнейшего анализа целесообразнее применять при оценке среднего метод экспоненциального сглаживания

$$x_{(n+1)j}^{\text{ср}} = \sum_{k=1}^n h_k(\alpha_j) x_{kj}, \quad j=\overline{1, m}, \quad (3)$$

где весовые коэффициенты h_k определяются через неизвестные пока α_j по формулам

$$h_k(\alpha_j) = (1 - \alpha_j)^{n-k} \alpha_j^{k-1} C_{n-1}^{k-1}, \quad C_{n-1}^{k-1} = \frac{(n-1)!}{(k-1)!(n-k)!},$$

$$\sum_{k=1}^n h_k(\alpha_j) = 1, \quad 0 \leq \alpha_j \leq 1, \quad j=\overline{1, m}.$$

При этом трендовое приращение $\Delta x_{(n+1)j}$ необходимо так же поставить в зависимость от коэффициентов h_k . Это позволит составить алгоритм, позволяющего на основе некоторого принципа согласованности изменения данных оценить прогнозные значения по всем показателям X_j . Обсуждение принципа и содержания алгоритма не является целью данной статьи и, поэтому, будем предполагать, что оценки по формулам (3) и (2) были сделаны.

Известно, что при любых оценках прогнозных значений нет 100% гарантии их наблюдения в будущем. Поэтому важным является задача определения для них доверительного интервала. Для адекватной оценки подобных интервалов может оказаться полезным метод главных компонент.

Модификация компонентного анализа

Многомерный анализ обычно проводится на основе предварительной стандартизации данных. При этом используются выборочная средняя и дисперсия, рассчитанных на основе средних арифметических имеющихся данных по показателям. Данные как бы считаются равноценными.

Допустим, что стандартизация проводится по средневзвешенным оценкам показателей X_j с весами $p_k, k=1, n+1$, т.е.

$$\bar{x}_j = \sum_{k=1}^{n+1} p_k x_{kj}, \quad s_j = \sqrt{\sum_{k=1}^{n+1} p_k (x_{kj} - \bar{x}_j)^2}, \quad \sum_{k=1}^{n+1} p_k = 1, \quad j=\overline{1, m}.$$

При этом прогнозные значения матрицы (1) входят в набор стандартизированных данных, где индекс «пр» уже не выделяется. Весовые

коэффициенты по всем показателям одинаковые, что будет отвечать единой стратегии статистических оценок. Коэффициенты корреляции между показателями следует тогда вычислять по формуле

$$r_{ij} = \frac{K(X_i, X_j)}{\sigma[X_i]\sigma[X_j]} \approx \frac{\sum_{k=1}^{n+1} p_k (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{s_i s_j}, \quad i, j = \overline{1, m}. \quad (4)$$

Пусть матрица $\mathbf{Z}_{(n+1) \times m}$ является результатом стандартизации данных матрицы (1). При этом корреляционная матрица есть произведение

$$\mathbf{R}_{m \times m} = \mathbf{Z}_{(n+1) \times m}^{\text{TP}} \mathbf{Z}_{(n+1) \times m}.$$

Тогда формулу (4) для оценки элементов корреляционной матрицы следует записать в виде

$$r_{ij} = \sum_{k=1}^{n+1} [\sqrt{p_k} \frac{(x_{ki} - \bar{x}_i)}{s_i}] [\sqrt{p_k} \frac{(x_{kj} - \bar{x}_j)}{s_j}], \quad i, j = \overline{1, m}. \quad (5)$$

Следовательно, элементы стандартизированной матрицы необходимо вычислять по формуле

$$z_{ij} = \sqrt{p_k} \frac{(x_{ij} - \bar{x}_j)}{s_j}, \quad i = \overline{1, n+1}, \quad j = \overline{1, m}. \quad (6)$$

При этом будут выполняться следующие соотношения

$$\sum_{k=1}^{n+1} \sqrt{p_k} z_{kj} = \sum_{k=1}^{n+1} p_k \frac{x_{kj} - \bar{x}_j}{s_j} = 0, \quad j = \overline{1, m}, \quad (7)$$

$$\sum_{k=1}^{n+1} z_{kj}^2 = \sum_{k=1}^{n+1} p_k \frac{(x_{kj} - \bar{x}_j)^2}{s_j^2} = 1, \quad j = \overline{1, m}. \quad (8)$$

Для проведения компонентного анализа можно было бы воспользоваться ПО STATISTICA, но оно ориентировано на вычисление средневзвешенных как средние арифметические, что отлично от предлагаемого подхода. Поэтому придется повторить все этапы факторного анализа, заложенного в этом программном обеспечении.

1) Определяются коэффициенты характеристического уравнения, корни которого λ_q , $1 \leq q \leq m$ являются собственными числами корреляционной матрицы $\mathbf{R}_{m \times m}$ с элементами, рассчитанными по формуле (5). Если отдельные показатели будут линейной комбинацией других, то q будет строго меньше m . Однако, в любом случае $\sum_q \lambda_q = m$. Простым и эффективным методом оценки этих коэффициентов является метод Фаддеева [23].

2) Из решения системы

$$(\mathbf{R}_{m \times m} - \lambda_q \cdot \mathbf{E}_{m \times m}) \cdot \mathbf{U}_{1 \times m} = \mathbf{0}_{1 \times m}$$

находятся собственные вектора $\bar{\mathbf{u}}_q = \mathbf{U}_{1 \times m}^{\text{TP}}$ корреляционной матрицы, отвечающие собственным значениям λ_q . Для однозначного решения системы q -ый компонент собственного вектора берется равным единице.

3) Определяются нормированные собственные вектора

$$\bar{\mathbf{v}}_q = \bar{\mathbf{u}}_q / |\bar{\mathbf{u}}_q| = (v_{1q}, \dots, v_{mq})^{\text{TP}}, \quad 1 \leq q \leq m.$$

4) Составляется матрица факторного отображения $\mathbf{A}_{m \times q}$, элементы которой определяются компонентами нормированных собственных векторов и соответствующих собственным значениям

$$a_{jq} = \sqrt{\lambda_q} \cdot v_{jq}, \quad j = \overline{1, m}, \quad 1 \leq q \leq m. \quad (9)$$

Данные коэффициенты отражают связь стандартизированных показателей Z_j с факторами (компонентами) F_q . В матричном виде она имеет вид

$$\mathbf{Z}_{(n+1) \times m}^{\text{TP}} = \mathbf{A}_{m \times q} \cdot \mathbf{F}_{q \times (n+1)}. \quad (10)$$

Согласно теореме Тэрстоуна [24], являющейся основной для факторного анализа, матрица факторного отображения обладает следующими свойствами

$$\mathbf{A}_{m \times q} \cdot \mathbf{A}_{m \times q}^{\text{TP}} = \mathbf{R}_{m \times m}, \quad \mathbf{A}_{m \times q}^{\text{TP}} \cdot \mathbf{A}_{m \times q} = \mathbf{\Lambda}_{q \times q}, \quad (11)$$

где у матрицы $\mathbf{\Lambda}_{q \times q}$ по главной диагонали стоят числа, обратные собственным значениям, а остальные равны нулю. Это позволит по формулы (10) определить факторную матрицу

$$\mathbf{F}_{q \times (n+1)} = \mathbf{\Lambda}_{q \times q}^{-1} \cdot \mathbf{A}_{m \times q}^{\text{TP}} \cdot \mathbf{Z}_{(n+1) \times m}^{\text{TP}}.$$

Следовательно, коэффициенты данной матрицы будут рассчитываться по формуле

$$f_{qk} = \frac{1}{\lambda_q} \sum_{j=1}^m z_{kj} a_{jq}, \quad 1 \leq q \leq m, \quad k = \overline{1, n+1}. \quad (12)$$

Не трудно убедиться, что в соответствии с (7) и (8) будут выполняться соотношения

$$\sum_{k=1}^{n+1} \sqrt{p_k} f_{qk} = 0, \quad \sum_{k=1}^{n+1} f_{kj}^2 = 1, \quad 1 \leq q \leq m. \quad (13)$$

При этом, как и для традиционного метода главных компонент будет выполняться равенство

$$\mathbf{F}_{q \times (n+1)} \cdot \mathbf{F}_{q \times (n+1)}^{\text{TP}} = \mathbf{E}_{q \times q}, \quad (14)$$

т.е. факторы будут ортогональны. Условие (14), где $\mathbf{E}_{q \times q}$ единичная матрица, обеспечивает выполнение первого соотношения в (11).

Веса p_k можно определять по-разному. Однако, для удобства пересчетов лучше их определять по типу экспоненциального сглаживания

$$p_k(\delta) = (1 - \delta)^{n+1-k} \delta^{k-1} C_n^{k-1}, \quad (15)$$

где

$$C_n^{k-1} = \frac{n!}{(k-1)!(n+1-k)!}, \quad 0 \leq \delta \leq 1, \quad k = \overline{1, n+1}.$$

За счет выбора δ можно реализовывать различные стратегии прогнозирования. Если δ ближе к 1, то с большим весом будут учитываться стандартизированные данные ближе к прогнозу

Сравнение традиционного и модифицированного методов главных компонент (МГК)

Данные			Традиционный МГК			Модифицированный МГК						
X_1	X_2	X_3	a_{iq}	F_1	F_2	p_k	Z_1	Z_2	Z_3	a_{iq}	F_1	F_2
2	5	-1	Z_1	0,92	-0,39	0,026	-0,35	0,48	0,45	Z_1	0,89	-0,45
4	3	-3	Z_2	-1,00	0,07	0,154	0,01	-0,33	-0,84	Z_2	-1,00	0,07
3	4	-2	Z_3	-0,87	-0,49	0,346	-0,65	0,64	0,20	Z_3	-0,58	-0,81
5	3	-2	λ_q	2,608	0,392	0,346	0,66	-0,50	0,20	λ_q	2,132	0,868
4,30	3,31	-2,07		87%	13%	0,130	0,12	-0,09	0,06		71%	29%

ному значению, а при значениях δ близких к нулю – дальнее. При $\delta = (n - 1)/(n + 1)$ больший вес будет у двух ближайших к прогнозному значению.

Демонстрационные сравнительные расчеты

При сравнительных расчетах предполагается, что динамика состояния исследуемого объекта характеризуется тремя показателями за четыре момента времени, значения которых представлены в первой части таблицы. При этом в пятой строке приведены возможные прогнозные значения показателей, рассчитанные по технологии, обозначенной формулами (2) и (3). Третий показатель имеет явную линейную зависимость от двух первых

$$X_3 = -13 + X_1 + 2X_2 \cdot X_3,$$

что находит отражение и в оценках прогнозных значений.

Во второй части таблицы представлены результаты расчетов, проведенных традиционным методом главных компонент с использованием ПО STATISTICA. Так как определитель корреляционной матрицы равен нулю, то факторов (компонент) F_q меньше чем стандартизованных показателей Z_j и матрица факторного отображения $A_{m \times q}$ прямоугольная. Собственные значения корреляционной матрицы λ_q указывают на то, что первая компонента будет главной, так как почти на 87% отражает вариацию значений исходных показателей.

В третьей части таблицы представлены результаты расчетов, проведенных изложенным выше модифицированным методом главных компонент. Первоначально по формуле (15) были определены веса значимости значений показателей, рассчитанных для $\delta = (4 - 1)/(5 + 1) = 0,6$. При этом два ближайших к прогнозному значению каждого показателя имеют наибольший вес. Стандартизация первоначальных данных, проводилась по формуле (6). Для определения матрицы факторного отображения методом Фаддеева были рассчитаны коэффициенты характеристического уравнения корреляционной матрицы, которое в итоге имеет вид

$$\lambda^2 - 3\lambda + 1,851 = 0.$$

Корни данного уравнения указывают на то, что первая компонента будет главной, но она менее информативна по сравнению с традиционным МГК. Однако это не есть недостаток, так как для построения комплексной оценки динамики состояния исследуемого объекта могут быть использованы все компоненты [25]. Кроме того, ниже в заключении будет отмечено, что значимость модифицированного подхода не в информативности компонент.

По известным λ_q можно рассчитать собственные вектора корреляционной матрицы и согласно формуле (9) элементы a_{iq} матрицы факторного отображения. Это позволит по формуле (12) оценить элементы факторной матрицы

$$F_{q \times (n+1)} = \begin{pmatrix} -0,498 & 0,386 & -0,624 & 0,455 & 0,077 \\ -0,202 & 0,759 & 0,198 & -0,573 & -0,124 \end{pmatrix}.$$

Не трудно проверить, что соотношения (13) и (14) будут выполняться. Матрица факторных отображений $A_{m \times q}$ будет удовлетворять теореме Тэрстоуна, выраженной матричными соотношениями (11).

Заключение

Используя формулу (10), справедливую как для традиционного, так и для представленного модифицированного метода главных компонент, мы снова вернемся к первоначальным значениям показателей, составляющих матрицу (1). Возникает вопрос – Для чего все приведенные выше рассуждения?

Используя формулы (6) и (12) отдельно рассмотрим элементы факторной матрицы на прогнозный $(n + 1)$ момент времени

$$f_{q(n+1)} = \frac{1}{\lambda_q} \sum_{j=1}^m a_{jq} \sqrt{p_{n+1}} \frac{(x_{(n+1)j} - \bar{x}_j)}{s_j}, 1 \leq q \leq m.$$

Прогнозные значения показателей $\Delta x_{(n+1)j}$ можно заменить выражением (2), где среднее оценивается по формуле (3). Проведя эти подстановки, после ряда преобразования окончательно можно записать

$$f_{q(n+1)} = \sum_{k=1}^n h_k(\alpha) \cdot \left[\sqrt{\frac{p_{n+1}(\delta)}{p_k(\delta)}} f_{qk} \right] + \Delta_{q(n+1)}, 1 \leq q \leq m,$$

где

$$\Delta_{q(n+1)} = \frac{\sqrt{p_{n+1}(\delta)}}{\lambda_q} \sum_{j=1}^m a_{jq} \frac{\Delta x_{(n+1)j}}{s_j}.$$

При этом полагается, что в формуле (3) $\alpha = \alpha_1 = \dots = \alpha_m$, т.е. согласованность изменения данных происходит при некоторых одинаковых α . Предварительный анализ показывает, что это условие будет выполняться.

Тогда можно утверждать, что значение факторов в прогнозный момент времени оценивается через средневзвешенную предыдущих их значений. Это позволит на $(n + 1)$ момент времени оценить дисперсию возможных значений факторов, а, следовательно, и доверительный интервал прогнозных значений. Данный доверительный интервал за счет

весов h_k будет отслеживать согласованность изменения данных, а за счет весов p_k стратегию прогнозирования. Вес $p_{(n+1)}$ имеет небольшое значение, так как веса p_k используются не для уточнения прогнозных значений показателей, а для уточнения доверительных интервалов для них.

За счет присутствия в процедуре прогнозирования метода главных компонент стратегия латентно будет учитывать взаимное влияние исходных показателей. Оценка доверительного интервала для прогнозных значений факторов позволит определить доверительные интервалы для прогнозных значений каждого отдельного показателя, а также некоторого показателя комплексной оценки динамики состояния исследуемого объекта.

Литература

1. Зайцев С.А., Сысоев А.П. Прогнозирование скоростной модели по данным ГИС // Технологии сейсморазведки. 2016. № 4. С. 56–60.
2. Gan S.O., Ahmad S. Multiple Linear Regression to Forecast Balance of Trade // Journal of Fundamental Sciences. 2011. Т. 7. № 2. С. 150–155.
3. Никитин В.В., Данилов И.П., Назаров А.А., Бобин Д.В. Комплексная модельная оценка финансового состояния предприятия // Региональная экономика: теория и практика. 2018. Т. 16. № 3. С. 551–566.
4. Хеннан Э. Многомерные временные ряды: Пер. с англ. М.: Издательство «МИР», 1974. 576 с.
5. Wang L. Advanced Multivariate Time Series Forecasting Models // Journal of Mathematics and Statistics. 2018. Т. 14. С. 253–260.
6. Борисов В.В., Луферов В.С. Метод многомерного анализа и прогнозирования состояния сложных систем и процессов на основе нечетких когнитивных темпоральных моделей // Системы управления, связи и безопасности. 2020. № 2. С. 1–23.
7. Матвеев М.Г., Сирота Е.А. Комбинированные модели нестационарных временных рядов с изменяющимися состояниями // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2016. № 3. С. 50–81.
8. Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space // Philosophical Magazine. 1901. Т. 2. № 6. С. 559–572.
9. Gorban A., Kegl B., Wunsch D., Zinovyev A. ed. Principal Manifolds for Data Visualisation and Dimension Reduction. LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007. 330 с.
10. Галиаскаров М.Р., Русинов Л.А. Система диагностики на базе нелинейного метода главных компонент и дискриминантного анализа Фишера // Известия СПбГТИ(ТУ). 2016. № 33. С. 91–96.
11. Belyavskiy G.I., Puchkov E.V. Nonlinear Principal Component Analysis Approach to Pattern Recognition // Modeling of Artificial Intelligence. 2016. Т. 9. № 1. С. 24–32.
12. Сидоренко А.В., Шакинко И.В. Модифицированный метод главных компонент при шифровании изображений с использованием динамического хаоса // Вестник БГУ. Серия 1: Физика. Математика. Информатика. 2014. № 3. С. 25–29.
13. Монахова Ю.Б., Цикин А.М., Муштакова С.П. Метод независимых компонент как альтернатива методу главных компонент и дискриминантным алгоритмам в обработке спектрометрических данных // Журнал аналитической химии. 2015. Т. 70. № 9. С. 1055–1061.
14. Kairov U., Cantini L., Greco A., Molkenov A., Czerwinska U., Barillot E., Zinovyev A. Determining the optimal number of independent components for reproducible transcriptomic data analysis // BMC Genomics. 2017. № 18(1). С. 7–12.
15. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы для экономистов и менеджеров: Учебник. М.: Финансы и статистика, 2003. 352 с.
16. Никитин В.В., Бобин Д.В., Назаров А.А. Рейтинг инвестиционного потенциала регионов Российской Федерации: многомерный статистический анализ // Экономика, статистика и информатика. Вестник УМО. 2014. № 3. С. 132–138.
17. Hinton G.E., Salakhutdinov R.R. Reducing the dimensionality of data with neural networks // Science. 2006. № 313(5786). С. 504–507.
18. Поляк Б.Т., Хлебников М.В. Метод главных компонент: робастные версии // Автоматика и телемеханика. 2017. № 3. С. 130–148.
19. Croux C., Haesbroeck G. Principal Component Analysis Based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies // Biometrika. 2000. Т. 87. № 3. С. 603–618.
20. Бурнаев Е.В., Чернова С.С. Об итеративном алгоритме подсчета взвешенных глав-

ных компонент // Информационные процессы. 2008. Т. 8. № 2. С. 99–107.

21. Delchambre L. Weighted principal component analysis: a weighted covariance eigendecomposition approach // Monthly Notices of the Royal Astronomical Society. 2014. № 446(2). С. 3545–3555.

22. Van Deun K., Thorrez L., Coccia M., Hasdemir D., Westerhuis J.A., Smilde A.K., Van Mechelen I. Weighted sparse principal component analysis // Chemometrics and Intelligent Laboratory Systems. 2019. Т. 195. С. 1–13. DOI: 10.1016/j.chemolab.2019.103875.

23. Гантмахер Ф.Р. Теория матриц. М.: Издательская фирма «Физико-математическая литература», 2010. 560 с.

24. Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике: Учебное пособие / Под ред. В.Н. Тамашевича. М.: ЮНИТИ-ДА-НА, 1999. 598 с.

25. Никитин В.В., Бобин Д.В., Назаров А.А. Алгоритм построения показателя комплексной оценки объекта исследования [Электрон. ресурс] // Проблемы и перспективы развития социально-экономического потенциала российских регионов: Материалы 8-й Всероссийской научно-практической конференции (13 декабря 2019г., Чебоксары, Россия). Чебоксары: Издательство Чувашского университета, 2019. С. 242–246. Режим доступа: <https://www.elibrary.ru/item.asp?id=41869567>.

References

1. Zaytsev S.A., Sysoyev A.P. Prediction of the Velocity Model Based on Well Logging Data. Tekhnologii seysmorazvedki = Seismic Technologies. 2016; 4: 56–60. (In Russ.)

2. Gan S.O., Ahmad S. Multiple Linear Regression to Forecast Balance of Trade. Journal of Fundamental Sciences. 2011; 7; 2: 150–155.

3. Nikitin V.V., Danilov I.P., Nazarov A.A., Bobin D.V. Complex model assessment of the financial condition of the enterprise. Regional'naya ekonomika: teoriya i praktika = Regional economy: theory and practice. 2018; 16; 3: 551–566. (In Russ.)

4. Kennan E. Mnogomernyye vremennyye ryady: Per. s angl. = Multidimensional time series: Tr. from Eng. Moscow: Publishing house «MIR»; 1974. 576 p. (In Russ.)

5. Wang L. Advanced Multivariate Time Series Forecasting Models. Journal of Mathematics and Statistics. 2018; 14: 253–260.

6. Borisov V.V., Luferov V.S. Method of multivariate analysis and forecasting of the state of complex systems and processes based on fuzzy cognitive temporal models. Sistemy upravleniya, svyazi i bezopasnosti = Control systems, communications and security. 2020; 2: 1–23. (In Russ.)

7. Matveyev M.G., Sirota Ye.A. Combined models of non-stationary time series with changing states. Vestnik VGU. Seriya: Sistemnyy analiz i informatsionnyye tekhnologii = Vestnik VSU. Series: System Analysis and Information Technology. 2016; 3: 50–81. (In Russ.)

8. Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine. 1901; 2; 6: 559–572.

9. Gorban A., Kegl B., Wunsch D., Zinovyev A. ed. Principal Manifolds for Data Visualisation and Dimension Reduction. LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007. 330 p.

10. Galiaskarov M.R., Rusinov L.A. Diagnostic system based on the nonlinear principal component

method and Fisher's discriminant analysis. Izvestiya SPbGTI(TU) = Izvestia SPbGTI (TU). 2016; 33: 91–96. (In Russ.)

11. Belyavskiy G.I., Puchkov E.V. Nonlinear Principal Component Analysis Approach to Pattern Recognition. Modeling of Artificial Intelligence. 2016; 9; 1: 24–32.

12. Sidorenko A.V., Shakinko I.V. Modified principal component method for image encryption using dynamic chaos. Vestnik BGU. Seriya 1: Fizika. Matematika. Informatika = Bulletin of BSU. Series 1: Physics. Mathematics. Computer science. 2014; 3: 25–29. (In Russ.)

13. Monakhova Yu.B., Tsikin A.M., Mushtakova S.P. The method of independent components as an alternative to the method of principal components and discriminant algorithms in the processing of spectrometric data. Zhurnal analiticheskoy khimii = Journal of Analytical Chemistry. 2015; 70; 9: 1055–1061. (In Russ.)

14. Kairov U., Cantini L., Greco A., Molkenov A., Czerwinska U., Barillot E., Zinovyev A. Determining the optimal number of independent components for reproducible transcriptomic data analysis. BMC Genomics. 2017; 18(1): 7–12.

15. Dubrov A.M., Mkhitarian V.S., Troshin L.I. Mnogomernyye statisticheskiye metody dlya ekonomistov i menedzherov: Uchebnik = Multivariate Statistical Methods for Economists and Managers: A Textbook. Moscow: Finance and Statistics; 2003. 352 p. (In Russ.)

16. Nikitin V.V., Bobin D.V., Nazarov A.A. Rating of the investment potential of the regions of the Russian Federation: multidimensional statistical analysis. Ekonomika, statistika i informatika. Vestnik UMO = Economics, statistics and informatics. Bulletin of UMO. 2014; 3: 132–138. (In Russ.)

17. Hinton G.E., Salakhutdinov R.R. Reducing the dimensionality of data with neural networks. Science. 2006; 313(5786): 504–507.

18. Polyak B.T., Khlebnikov M.V. Principal Component Method: Robust Versions. Avtomatika i

telemekhanika = Automation and Remote Control. 2017; 3: 130–148. (In Russ.)

19. Croux C., Haesbroeck G. Principal Component Analysis Based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. *Biometrika*. 2000; 87; 3: 603–618.

20. Burnayev Ye.V., Chernova S.C. On an iterative algorithm for calculating weighted principal components. *Informatsionnyye protsessy = Information processes*. 2008; 8; 2: 99–107. (In Russ.)

21. Delchambre L. Weighted principal component analysis: a weighted covariance eigendecomposition approach. *Monthly Notices of the Royal Astronomical Society*. 2014; 446(2): 3545–3555.

22. Van Deun K., Thorrez L., Coccia M., Hasdemir D., Westerhuis J.A., Smilde A.K., Van Mechelen I. Weighted sparse principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 2019; 195: 1–13. DOI: 10.1016/j.chemolab.2019.103875.

23. Gantmakher F.R. *Teoriya matrits = Matrix theory*. Moscow: Publishing firm «Physical and mathematical literature»; 2010. 560 p. (In Russ.)

24. Soshnikova L.A., Tamashevich V.N., Uyebe G., Shefer M. *Mnogomernyy statisticheskiy analiz v ekonomike: Uchebnoye posobiye / Pod red. V.N. Tamashevicha = Multivariate statistical analysis in economics: Textbook - Ed. V.N. Tamashevich*. Moscow: UNITI-DANA; 1999. 598 p. (In Russ.)

25. Nikitin V.V., Bobin D.V., Nazarov A.A. Algorithm for constructing an indicator for a comprehensive assessment of the object of research [Internet]. *Problemy i perspektivy razvitiya sotsial'no-ekonomicheskogo potentsiala rossiyskikh regionov: Materialy 8-y Vserossiyskoy nauchno-prakticheskoy konferentsii = Problems and prospects for the development of the socio-economic potential of Russian regions: Materials of the 8th All-Russian Scientific and Practical Conference (December 13, 2019, Cheboksary, Russia)*. Cheboksary: Publishing House of the Chuvash University; 2019: 242–246. Available from: <https://www.elibrary.ru/item.asp?id=41869567>. (In Russ.)

Сведения об авторах

Виктор Васильевич Никитин

К.ф.-м.н., доцент, профессор кафедры актуарной и финансовой математики

Чувашский государственный университет

им. И.Н. Ульянова,

Чебоксары, Россия

Эл. почта: vvn22@yandex.ru

Дмитрий Витальевич Бобин

Старший преподаватель кафедры актуарной и финансовой математики

Чувашский государственный университет

им. И.Н. Ульянова, Чебоксары, Россия

Эл. почта: dimbobin@mail.ru

Information about the authors

Viktor V. Nikitin

Cand. Sci. (Physics and Mathematics), Associate Professor, Professor of the Chair of Actuarial and Financial Mathematics

Chuvash State University named after I.N. Ulyanov, Cheboksary, Russia

E-mail: vvn22@yandex.ru

Dmitriy V. Bobin

Senior lecturer of the Chair of Actuarial and Financial Mathematics

Chuvash State University named after I.N. Ulyanov, Cheboksary, Russia

E-mail: dimbobin@mail.ru