

УДК 33

Применение моделей сжатия размерности при выявлении скрытых экономических закономерностей

Хамидуллин Равгат Явдатович

Кандидат технических наук,
доцент,

завкафедрой высшей математики и естественнонаучных дисциплин,
Московский финансово-промышленный университет «Синергия»,
125190, Российская Федерация, Москва, просп. Ленинградский, 80;
e-mail: ravgat@yandex.ru

Аннотация

В настоящее время в экономической науке в целом и финансах в частности являются доказанными незначительное количество неортодоксальных статистических значимых закономерностей, характерных для стран со всеми типами экономик. К таким зависимостям, в частности, относятся кривая Филипса. В работе был проведен анализ макрофинансовых показателей на основании которого можно сделать вывод о том, что даже самый простой анализ некоррелируемых показателей макрофинансовой статистики указывает на то, что между ними существуют сложные взаимозависимости, изучение которых может выявить скрытые механизмы исследуемого явления. В работе показано, что применение инструментов признакового пространства позволяет осуществлять анализ скрытых взаимосвязей между явлениями в большей степени, чем распространенный в научной литературе линейный и множественный корреляционно-регрессионный анализ. При этом показано, что графические инструменты определения правила остановки выявления факторов является достоверным инструментом при исследовании сложных социально-экономических явлений и процессов.

Для цитирования в научных исследованиях

Хамидуллин Р.Я. Применение моделей сжатия размерности при выявлении скрытых экономических закономерностей // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 11А. С. 234-242.

Ключевые слова

Методы сжатия признакового пространства, статистический анализ, макрофинансовая статистика, статистические зависимости, метод главных компонент.

Введение

В настоящее время происходит все большее увеличение информации, создаются и развиваются технологии сбора, обработки, анализа и хранения больших данных. Так, например, показатели развития банковской системы или макроэкономической статистики составляют несколько сотен показателей, очевидно, что анализ данных показателей посредством распространенных статистических инструментов, таких как инструменты многомерной классификации и корреляционно-регрессионный анализ.

В этих условиях становится все более актуальной проблема поиска экономико-математических инструментов, позволяющих получать объективную информацию о происходящих процессах, выявлять сложные закономерности развития социально-экономических систем. В ряду таких методов особое место занимают методы сжатия признаков пространства, позволяющие заменять исходные данные признаков пространства, описывающего социально-экономические явления или процессы на условные показатели, при этом показателем, характеризующим качество сжатия, является уровень объясненной дисперсии.

Следовательно, применение инструментов статистических анализов данных представляет собой значимую область исследований, расширяющую представления о возможностях применения различных инструментов при анализе закономерностей социально-экономического развития.

Применение инструментов статистического анализа с целью сжатия размерности

Ключевыми целями факторного анализа является редукция данных с целью определения более глубоких взаимосвязей между переменными. В настоящее время методология факторного анализа широко применяется в современных исследованиях.

Идея данного метода заключается в следующем.

В ряде случаев, при описании некоторого явления или процесса используются показатели со схожими характеристиками.

Очевидно, что для первичного определения мультиколлинеарности получаемого признаков пространства используются показатели корреляции, на их основании можно исключить их признаков пространства показатели с наиболее высокими значениями корреляции. Однако, при использовании такого метода может возникнуть ситуация, когда даже незначительные расхождения между показателями, могут оказывать влияние на характеристики исследуемого явления.

Посредством применения метода главных компонент применяются инструменты, согласно которым два наиболее близких фактора заменяются на их линейную комбинацию. Однако в случае, когда количество факторов значительное количество применение данного инструмента затруднительно.

В случае, когда признаков пространство составляет значительное количество факторов применяются инструменты выделения главных компонент. Суть метода заключается в применении процедуры выделения главных компонент представляет собой в самом общем виде

вращения, максимизирующего дисперсию исходного пространства переменных. Основная идея заключается в том, что на графике поля корреляции можно повернуть ось X так, чтобы она совпала с прямой регрессией. В результате применения данного метода получаются значения условного фактора, для которого критерием является максимизация дисперсии и минимизации разброса вокруг нее.

После того, как находится линия, для которой дисперсия максимальна, остается некоторый разброс данных, которые повторно используются посредством метода вращения. Следовательно, в результате работы последовательно выделяются факторы в той или иной степени описывающие характеристики явления или процесса, так как каждый последующий фактор определяется так, чтобы максимизировать вариацию, оставшуюся от предыдущих. При этом все данные факторы являются независимыми друг от друга, некоррелированными и ортогональными.

Необходимо отметить, что решение относительно того, сколько факторов необходимо выявлять подобным образом полностью зависит от того, какой уровень объясненной дисперсии является достаточным для целей исследования. Так, очевидно, что 100% дисперсии может быть выделено только посредством аналогичного исходной выборке количества факторов.

В частности, для определения необходимого уровня дисперсии используются следующие инструменты:

- критерий Кайзера, согласно которой отбираются только такие факторы, у которых уровень объясненной дисперсии составляет больше определенного значения;
- критерий «каменистой осыпи» является графическим методом, согласно которому определяется место на графике, где убывание собственных значений максимально замедляется. В частности, предполагается, что справа от этой точки находится только «факториальная осыпь», уровень которой является термином, применяемым в геологии.

В результате применения метода сжатия пространства, исходное признаковое пространство заменяется на условные факторы в случае, если условные показатели могут существенно отличаться от исходных значений, при этом наличие ортогональных исходным факторам указывает на наличие скрытых зависимостей между показателями.

Апробация метода сжатия признакового пространства на примере показателей макрофинансовой статистики

С целью апробации метода сокращения признакового пространства были использованы следующие показатели Центрального Банка, наиболее системно характеризующие макрофинансовое состояние российской экономики:

- Изменение наличных денег в обращении (вне Банка России).
- Изменение остатков средств на счетах расширенного правительства в Банке России и прочие операции.
- Регулирование Банком России обязательных резервов кредитных организаций.
- Операции Банка России на внутреннем валютном рынке.
- Нетто-объем операций Банка России по предоставлению и абсорбированию ликвидности.
- Стоимость бивалютной корзины.

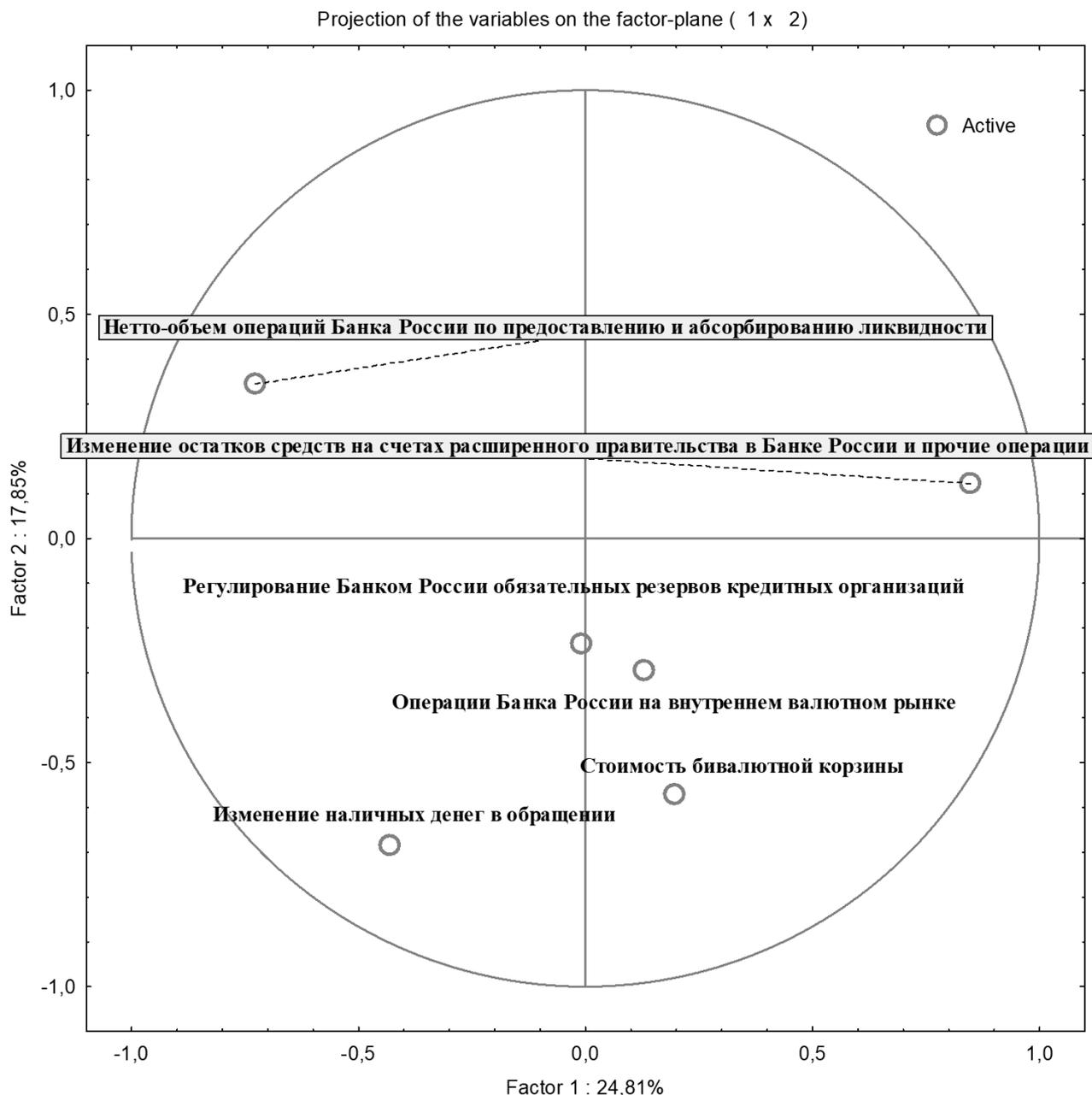
На основании исследования показателей линейной корреляции было выявлено, что между данными показателями нет прямой или обратной статистической зависимости.

Таблица 1 – Корреляция между показателями макрофинансовой статистики

Переменные	Изменение наличных денег в обращении (вне Банка России)	Изменение остатков средств на счетах расширенного правительства в Банке России и прочие операции	Регулирование Банком России обязательных резервов кредитных организаций	Операции Банка России на внутреннем валютном рынке	Нетто-объем операций Банка России по предоставлению и абсорбированию ликвидности	Стоимость бивалютной корзины
Изменение наличных денег в обращении (вне Банка России)	1,000000	-0,261969	0,005122	-0,0217	-0,010938	0,031671
Изменение остатков средств на счетах расширенного правительства в Банке России и прочие операции	-0,261969	1,000000	-0,012037	-0,01312	-0,396687	0,071795
Регулирование Банком России обязательных резервов кредитных организаций	0,005122	-0,012037	1,000000	0,005789	-0,006368	0,013814
Операции Банка России на внутреннем валютном рынке	-0,021695	-0,013115	0,005789	1,000000	-0,087817	0,004448
Нетто-объем операций Банка России по предоставлению и абсорбированию ликвидности	-0,010938	-0,396687	-0,006368	-0,08782	1,000000	-0,066582
Стоимость бивалютной корзины	0,031671	0,071795	0,013814	0,004448	-0,066582	1,000000

[Источник данных: Центральный банк, расчеты автора]

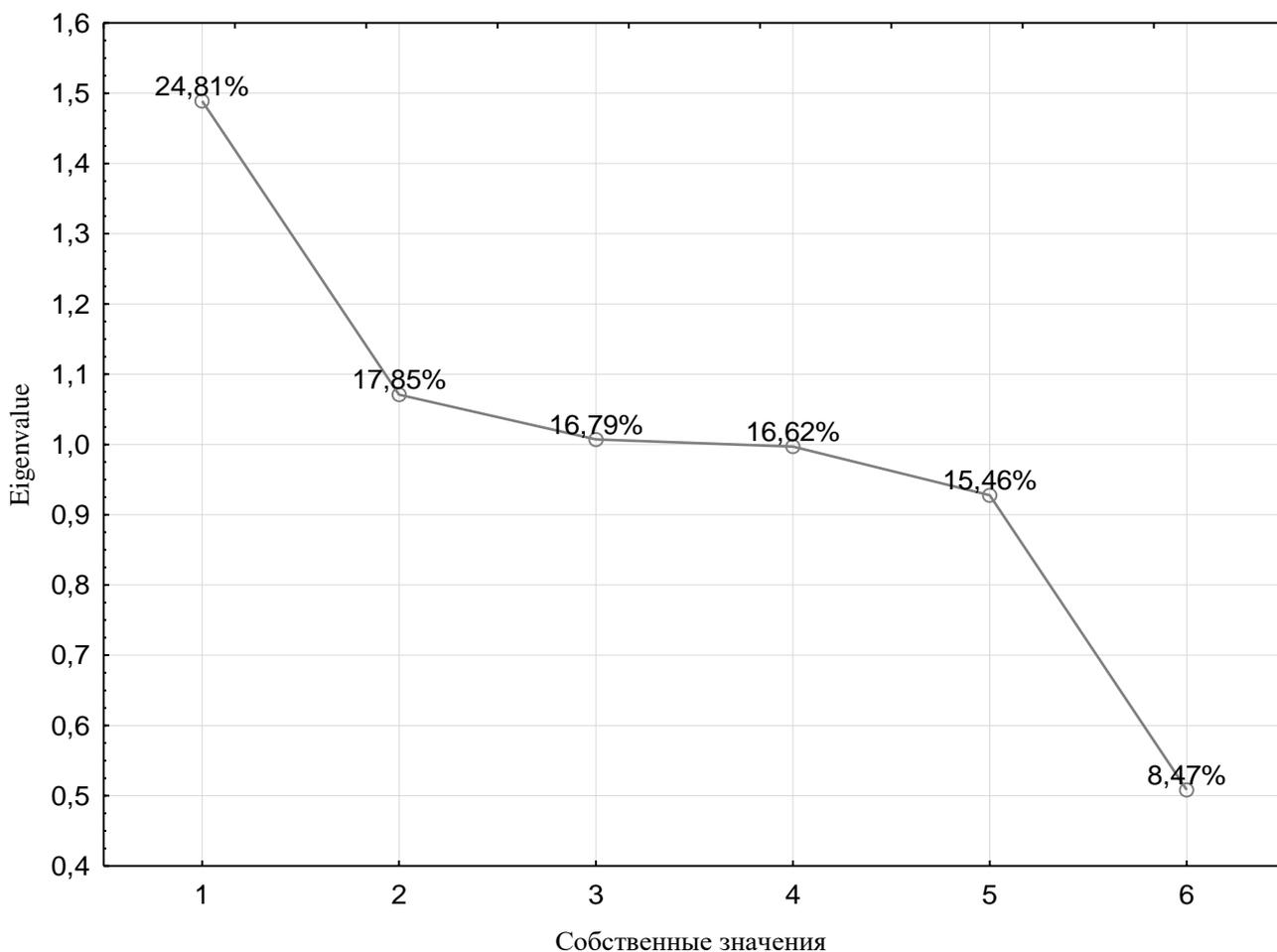
В результате применения инструментов применялся метод главных компонент, в результате были определены уровни объясненной дисперсии условных факторов.



[Источник данных: Центральный банк, расчеты автора]

Рисунок 1 – Проекция характеристик факторов

Необходимо отметить, что проекция характеристик факторов для первых двух факторов указывает на то, что из 6 показателей 4 в наибольшей степени характеризуются именно первыми двумя факторами. Таким образом, между некоррелируемыми факторами существуют более сложные зависимости, характер которых необходимо дополнительно изучать посредством апробации ряда исследовательских гипотез.



[Источник данных: Центральный банк, расчеты автора]

Рисунок 2 – Значение объясненной дисперсии показателей

Как видно из полученных результатов, максимальный уровень дисперсии имеет фактор с уровнем объясненной дисперсии 24,81%. Согласно критерию «каменистой осыпи», 6 факторов можно заменить на первые 4 фактора, при этом совокупный уровень объясненной дисперсии составляет 76,07%. Следовательно, совокупность факторов макрофинансовой статистики может быть описано с высокой степенью выявленной дисперсии посредством незначительного количества показателей. При этом очевидно, что необходимо осуществлять поиск новых показателей, описывающих зависимости между данными показателями.

Заключение

В настоящее время в экономической науке в целом и финансах в частности являются доказанным незначительное количество неортодоксальных статистических значимых закономерностей, характерных для всех экономических условий. Однако в то же время, даже самый простой анализ некоррелируемых показателей макрофинансовой статистики указывает на то, что между ними существуют сложные взаимозависимости, изучение которых может выявить скрытые механизмы исследуемого явления.

Таким образом, применение инструментов признакового пространства позволяет осуществлять анализ скрытых взаимосвязей между явлениями в большей степени, чем распространенный в научной литературе линейный и множественный корреляционно-регрессионный анализ. При этом графические инструменты определения правила остановки выявления факторов является достоверным инструментом при исследовании сложных социально-экономических явлений и процессов.

Библиография

1. Гулиян Б.Ш., Хамидуллин Р.Я. Выбор формы налогообложения с учетом зависимости величины чистой прибыли от квалификации производственного персонала // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 6А. С. 184-189.
2. Гулиян Б.Ш., Хамидуллин Р.Я. Оптимальное распределение инвестиций для достижения максимальной величины основных фондов малого предприятия с учетом квалификации производственного персонала // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 5А. С. 204-209.
3. Гулиян Б.Ш., Хамидуллин Р.Я. Решение задачи оптимизации инвестиций в основные фонды малого предприятия с учетом квалификации персонала при выборе формы налогообложения // Экономика: вчера, сегодня, завтра. 2018. Том 8. № 3А. С. 285-290.
4. Константинова Е.В. Глобализационные факторы становления сферы финансовых технологий // Экономика: вчера, сегодня, завтра. 2018. №9. С. 248-261.
5. Константинова Е.В. Обеспечение безопасности данных при использовании технологии блокчейн в кластерных и облачных вычислениях // Инновации и инвестиции. 2018. № 12. С. 340-347.
6. Константинова Е.В. Применение нейронных сетей для целей оптимизации состояний биржевой торговли // Инновации и инвестиции. 2018. № 7. С. 254-259.
7. Claveria O., Poluzzi A. Positioning and clustering of the world's top tourist destinations by means of dimensionality reduction techniques for categorical data // Journal of Destination Marketing & Management. 2017. V. 6. № 1. P. 22-32.
8. Cunningham J.P., Ghahramani Z. Linear dimensionality reduction: Survey, insights, and generalizations // The Journal of Machine Learning Research. 2015. V. 16. №1. P. 2859-2900.
9. Ding R. et al. Yading: fast clustering of large-scale time series data // Proceedings of the VLDB Endowment. 2015. V. 8. №5. P. 473-484.
10. Gisbrecht A., Hammer B. Data visualization by nonlinear dimensionality reduction // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2015. V. 5. №2. P. 51-73.
11. Moore J.H., Andrews P.C. Epistasis analysis using multifactor dimensionality reduction // Epistasis. Humana Press, New York, NY, 2015. P. 301-314.
12. Shen W. et al. Multi-scale convolutional neural networks for lung nodule classification // International Conference on Information Processing in Medical Imaging. Springer, Cham, 2015. P. 588-599.
13. Stahnke J. et al. Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions // IEEE transactions on visualization and computer graphics. 2016. V. 22. №1. P. 629-638.

The use of compression models of dimension in identifying hidden economic patterns

Ravgat Ya. Khamidullin

PhD in Technical Science, Associate Professor,
Head of the Department of Higher Mathematics and Natural Sciences,
Moscow Finance and Industry University "Synergy",
125190, 80, Leningradskii av., Moscow, Russian Federation;
e-mail: ravgat@yandex.ru

Ravgat Ya. Khamidullin

Abstract

At present, in economics in general and in finance in particular, a small number of unorthodox statistically significant patterns characteristic of countries with all types of economies are proven. These dependencies, in particular, include the Phillips curve. The paper analyzed the macro-financial indicator based on which we can conclude that even the simplest analysis of uncorrelated indicators of macro-financial statistics indicates that there are complex interdependencies between them, the study of which can reveal the hidden mechanisms of the phenomenon under study. The paper shows that the use of tools of attribute space allows for the analysis of hidden relationships between the phenomena to a greater extent than is common in the linear scientific literature and multiple correlation and regression analysis. It is shown that graphical tools for determining the rule for stopping the identification of factors is a reliable tool in the study of complex socio-economic phenomena and processes. The use of attribute space tools allows for the analysis of hidden relationships between phenomena to a greater extent than linear and multiple correlation and regression analysis common in the scientific literature. At the same time, graphic tools for determining the rule for stopping the identification of factors are a reliable tool in the study of complex socio-economic phenomena and processes.

For citation

Khamidullin R.Ya. (2018) Primenenie modelei szhatiya razmernosti pri vyyavlenii skrytykh ekonomicheskikh zakonomernostei [The use of compression models of dimension in identifying hidden economic patterns]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (11A), pp. 234-242.

Keywords

Methods of compression of the attribute space, statistical analysis, macro-financial statistics, statistical dependencies, the method of principal components.

References

1. Claveria O., Poluzzi A. (2017) Positioning and clustering of the world's top tourist destinations by means of dimensionality reduction techniques for categorical data. *Journal of Destination Marketing & Management*, 6(1), pp. 22-32.
2. Cunningham J.P., Ghahramani Z. (2015) Linear dimensionality reduction: Survey, insights, and generalizations. *The Journal of Machine Learning Research*, 16(1), pp. 2859-2900.
3. Ding R. et al. (2015) Yading: fast clustering of large-scale time series data. *Proceedings of the VLDB Endowment*, 8 (5), pp. 473-484.
4. Gisbrecht A., Hammer B. (2015) Data visualization by nonlinear dimensionality reduction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(2), pp. 51-73.
5. Guliyani B.Sh., Khamidullin R.Ya. (2018) O voprosakh optimizatsii osnovnykh fondov malogo predpriyatiya s uchetom kvalifikatsii proizvodstvennogo personala pri vybere vtoroi formy nalogooblozheniya (chistaya pribyl') [Optimal investment allocation for maximum value of fixed assets of small enterprises based on production personnel qualification]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (5A), pp. 204-209.
6. Guliyani B.Sh., Khamidullin R.Ya. (2018) Reshenie zadachi optimizatsii investitsii v osnovnye fondy malogo predpriyatiya s uchetom kvalifikatsii personala pri vybere formy nalogooblozheniya po dokhodu [Solving the problem of optimization of investments in fixed assets of small enterprise based on staff qualification when choosing a form of taxation]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (3A), pp. 285-290.
7. Guliyani B.Sh., Khamidullin R.Ya. (2018) Vybory formy nalogooblozheniya s uchetom zavisimosti velichiny chistoi pribyli ot kvalifikatsii proizvodstvennogo personala [Choice of a form of taxation based on dependence of net profit on skills of manufacturing personnel]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 8 (6A), pp. 184-189.

8. Konstantinova E.V. (2018) Globalizatsionnye faktory stanovleniya sfery finansovykh tekhnologii [Globalization factors of the development of the sphere of financial technologies]. *Ekonomika: vchera, segodnya, zavtra* [Economics: Yesterday, Today and Tomorrow], 9, pp. 248-261.
9. Konstantinova E.V. (2018) Obespechenie bezopasnosti dannykh pri ispol'zovanii tekhnologii blokchein v klasternykh i oblachnykh vychisleniyakh [Ensuring data security when using blockchain technology in cluster and cloud computing]. *Innovatsii i investitsii* [Innovations and investments], 2, pp. 340-347.
10. Konstantinova E.V (2018) Primenenie neironnykh setei dlya tselei optimizatsii sostoyanii birzhevoi trgovli [The use of neural networks for the purpose of optimizing the state of exchange trading]. *Innovatsii i investitsii* [Innovations and investments], 7, pp. 254-259.
11. Moore J.H., Andrews P.C. (2015) Epistasis analysis using multifactor dimensionality reduction. In: *Epistasis*. Humana Press, New York, NY.
12. Shen W. et al. (2015) Multi-scale convolutional neural networks for lung nodule classification. In: *International Conference on Information Processing in Medical Imaging*. Springer, Cham.
13. Stahnke J. et al. (2016) Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions. *IEEE transactions on visualization and computer graphics*, 22(1), pp. 629-638.