

ПОСТРОЕНИЕ МОДЕЛЕЙ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ УРОВНЯ РИСКОВ БАНКОВСКОГО СЕКТОРА РЕСПУБЛИКИ БЕЛАРУСЬ НА ОСНОВЕ ИНТЕГРАЛЬНЫХ ПОКАЗАТЕЛЕЙ

В. О. СУВАЛОВ¹⁾, Е. А. МИНЮКОВИЧ²⁾

¹⁾Национальный банк Республики Беларусь, пр. Независимости, 20, 220008, г. Минск, Беларусь

²⁾Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

Исследуется процесс построения эконометрической модели для создания автоматизированного алгоритма прогнозирования уровня рисков в банковском секторе Республики Беларусь. Отмечено, что внутренние свойства алгоритма основаны на формализации процесса построения эконометрической модели с выполнением необходимых предпосылок, а также на машинном обучении.

Ключевые слова: модель; прогнозирование; риски; эконометрический инструментарий; интегральные показатели; банковский сектор.

Благодарность. Авторы выражают благодарность Петру Алексеевичу Мамановичу, заместителю председателя правления Национального банка Республики Беларусь, и Сергею Николаевичу Шевчуку, начальнику управления анализа рисков банковской системы Главного управления банковского надзора Национального банка Республики Беларусь.

CONSTRUCTION OF MODELS OF ANALYSIS AND FORECASTING OF THE LEVEL OF RISKS OF THE BANKING SECTOR OF THE REPUBLIC OF BELARUS ON THE BASIS OF INTEGRATED INDICATORS

V. O. SUVALAU^a, K. A. MINIUKOVICH^b

^aNational Bank of the Republic of Belarus, 20 Niezaliežnasci Avenue, Minsk 220008, Belarus

^bBelarusian State University, 4 Niezaliežnasci Avenue, Minsk 220030, Belarus

Corresponding author: K. A. Miniukovich (miniukovich@bsu.by)

The process of constructing an econometric model for the purposes of creating an automated algorithm for forecasting the level of risks in the banking sector of the Republic of Belarus is being explored. The internal properties of the algorithm are based on the formalization of the process of constructing an econometric model with the fulfillment of the prerequisites, as well as machine learning.

Образец цитирования:

Сувалов В. О., Минюкович Е. А. Построение моделей анализа и прогнозирования уровня рисков банковского сектора Республики Беларусь на основе интегральных показателей // Журн. Белорус. гос. ун-та. Экономика. 2018. № 1. С. 20–28.

For citation:

Suvalau V. O., Miniukovich K. A. Construction of models of analysis and forecasting of the level of risks of the banking sector of the Republic of Belarus on the basis of integrated indicators. *J. Belarus. State Univ. Econ.* 2018. No. 1. P. 20–28 (in Russ.).

Авторы:

Валентин Олегович Сувалов – специалист первой категории управления регулирования ликвидности Главного управления операций на финансовых рынках.

Екатерина Александровна Минюкович – кандидат экономических наук, доцент; доцент кафедры экономической информатики экономического факультета.

Authors:

Valentin O. Suvalau, specialist of the liquidity regulation department of the Financial Markets Operations Directorate.

valentin.suvalau@gmail.com

Katsiaryna A. Miniukovich, PhD (economics), docent; associate professor at the department of economic informatics, faculty of economics.

miniukovich@bsu.by

Key words: model; forecasting; risks; econometric tools; integrated indicators; banking sector.

Acknowledgments. The authors are very grateful to Petr A. Mamanovich, deputy Chairman of the Board, National Bank of the Republic of Belarus and Sergey N. Shevchuk, head of banking system risk analysis department of the Banking Supervision Main Directorate, National Bank of the Republic of Belarus.

Введение

Любое экономически активное предприятие (организация) вне зависимости от того, в реальном или финансовом секторе осуществляет оно свою деятельность, сталкивается с теми или иными видами риска. Банковский сектор Республики Беларусь, определяемый нами как совокупность всех банков, осуществляющих свою деятельность на территории страны, также подвержен данному явлению.

Вопросы анализа рисков и их влияния стали объектом научного изучения У. Бека, Н. Лумана, П. Коха, Г. А. Риттера, Н. Крокфорда, Р. Пабста и др. [1–6]. Также риски деятельности на финансовых рынках анализировались в работе А. А. Первозванского и Т. Н. Первозванской «Финансовый рынок: расчет и риск» [7].

В Республике Беларусь одним из ведущих специалистов в анализе инфляционных рисков страны является А. Г. Галова [8]. Она занимается проблемами взаимодействия финансового и реального секторов экономики, а также тенденциями влияния финансового сектора Республики Беларусь на деятельность и развитие реального сектора страны.

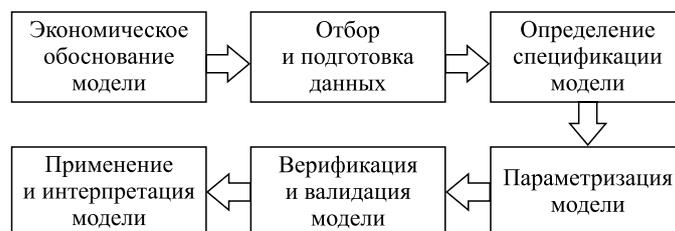
В данной статье представлен авторский подход к прогнозированию уровня рисков в банковском секторе и влияния рисков на целевые показатели развития банков путем применения эконометрического моделирования и современных алгоритмов машинного обучения. На базе указанного подхода построены две группы моделей. Первая позволяет на основе значений уровней четырех видов банковских рисков – кредитного, валютного, процентного и риска ликвидности – прогнозировать значения целевых показателей, в частности рентабельность собственного капитала и достаточность нормативного капитала. Вторая группа включает модели прогнозирования временных рядов для четырех названных выше видов банковских рисков. Предлагаемый подход к построению моделей может быть адаптирован для прогнозирования других целевых показателей, а также иных видов банковских рисков.

Разработанные модели положены в основу алгоритма, который реализован на языке программирования R в среде *RStudio*. Авторская программа¹ автоматизирует процесс прогнозирования уровня заданных пользователем банковских рисков и целевых показателей, а также визуализирует полученные результаты. Программа успешно применяется в Управлении анализа рисков банковского сектора Национального банка Республики Беларусь.

Теоретические основы построения моделей влияния уровней рисков на целевые показатели развития банков

Прежде чем приступить к разработке эконометрической модели для прогнозирования целевых показателей на основе значений уровней заданных банковских рисков, необходимо формализовать процесс построения эконометрической модели. По своей сути данный процесс не должен отличаться от этапов построения модели, через которые проходит аналитик. Классический порядок построения эконометрической модели может быть представлен в виде схемы (см. рисунок).

При обнаружении проблем на этапе верификации и валидации модели возможен возврат на этап определения спецификации.



Порядок построения эконометрической модели
The procedure for constructing an econometric model

¹Автором программы является В. О. Сувалов.

В целях обеспечения интерпретируемости модели было принято решение искать зависимость значений целевых показателей от уровней рисков в виде линейной регрессии. Последовательно рассмотрим каждый из этапов построения эконометрической модели линейной регрессии. При этом определим этапы, которые возможно полностью или частично автоматизировать.

Этап экономического обоснования является первым в процессе создания эконометрической модели и во многом определяет качество результатов моделирования. В первую очередь это связано с тем, что в данном процессе еще до определения самих исследуемых зависимостей устанавливаются цели моделирования. Именно от них зависят те требования, которые на следующих этапах будут предъявляться к модели. Ниже мы рассмотрим этот вопрос более детально.

На этапе экономического обоснования модели нельзя обойти стороной решение о границах доверительного интервала при проверке гипотез и анализе результатов, получаемых на следующих этапах. Конечно, данное решение может изменяться в процессе непосредственного анализа полученной модели. Более того, исследователи ориентируются на границы доверительного интервала, понимая, что имеют дело с вероятностями принятия той или иной гипотезы о свойствах, присущих модели. Во многом «строгость» при принятии той или иной гипотезы зависит от целей, которые ставятся перед моделью.

Так, для моделей, нацеленных на выявление степеней зависимости между показателями, требуется весьма строгое выполнение предпосылок теоремы Гаусса – Маркова, а именно отсутствие или коррекция мультиколлинеарности, автокорреляции и гетероскедастичности. В прогнозных и регулирующих моделях основное требование – отсутствие гетероскедастичности. Мультиколлинеарность может присутствовать в модели в качестве отражения действия синергетического эффекта, если от проблемы мультиколлинеарности действительно не удастся избавиться. Автокорреляция, в свою очередь, может отражать действие процессов, обладающих высокой степенью инерционности.

Далее идет **этап отбора и поиска данных для модели**, название которого полностью отражает его содержание.

Затем следует **этап спецификации модели**. Спецификация модели относится к определению того, какие независимые переменные должны быть включены в уравнение регрессии или исключены из него [9]. В целом спецификация регрессионной модели должна основываться главным образом на теоретических, а не эмпирических или методологических соображениях. Именно поэтому этап теоретического обоснования модели предшествует прочим. Модель множественной регрессии фактически является теоретическим утверждением о причинной связи между одной или несколькими независимыми переменными и зависимой переменной. При этом крайне важно правильно определить спецификацию модели, ведь наши оценки параметров модели и их интерпретация зависят от правильной спецификации модели, так как очевидно, что неправильная спецификация приведет к накоплению ошибок и на последующих этапах построения модели. Следовательно, проблемы могут возникать, когда мы неверно определяем модель, а конкретнее зависимость переменных модели. Существует два основных типа ошибок спецификации. Первый из них связан с ошибочным включением в уравнение регрессии независимой переменной, теоретически не имеющей значения. Ошибки второго типа обусловлены исключением из уравнения регрессии независимой переменной, теоретически релевантной [10, с. 172–197].

Для того чтобы удостовериться в отсутствии ошибок спецификации, можно применять тест Грейнджера. Он позволяет определять наличие причинно-следственных связей. Тест был разработан Клайвом Грейнджером, английским экономистом, лауреатом премии Шведского национального банка по экономическим наукам памяти Альфреда Нобеля 2003 г. за разработку методов анализа экономических временных рядов с общими трендами [11].

Тест Грейнджера представляет собой статистический тест, направленный на определение того, полезен ли один временной ряд для прогнозирования другого [12]. Обычно регрессии отражают простые корреляции, но К. Грейнджер утверждал, что причинность в экономике можно проверить, измеряя способность прогнозировать будущие значения временного ряда, используя предыдущие значения другого временного ряда. Поскольку вопрос об истинной причинности глубоко философский и из-за наличия идеи *post hoc ergo propter hoc* («после этого значит по причине этого») нетрудно ошибочно предположить, что одна вещь, предшествующая другой, может быть использована в качестве доказательства причинно-следственной связи, специалисты в области эконометрики утверждают, что тест Грейнджера находит только так называемую предсказательную причинность [13, р. 254].

Временной ряд X называется причиной по Грейнджеру по отношению к ряду Y , если может быть доказано, обычно через серию t -тестов и F -тестов с запаздывающими значениями X (и с включенными запаздывающими значениями Y), что эти значения X обеспечивают статистически значимую информацию о будущих значениях Y , т. е. по их изменениям можно предугадывать будущие изменения переменной Y .

Тест Грейнджера остается популярным методом анализа причинности во временных рядах из-за его вычислительной простоты. Первоначальный вариант, предложенный К. Грейнджером, не учитывает скрытые эффекты и не фиксирует мгновенные и нелинейные причинно-следственные связи, хотя к настоящему времени было предложено несколько расширений теста Грейнджера для решения этих проблем [14, p. 327–352].

Далее следует *этап параметризации модели*. Параметризация – это математический процесс, выражения состояния системы или модели как функции некоторых независимых величин, называемых параметрами. В нашем случае такими параметрами выступают коэффициенты при независимых переменных модели. Состояние системы обычно определяется конечным набором координат, а параметризация состоит, таким образом, из одной функции нескольких действительных переменных для каждой координаты. Количество параметров соответствует числу степеней свободы системы. Существует множество техник параметризации, например параметризация по Фейнману [15] или параметризация по Швингеру [16]. Однако решение этих вопросов лежит, скорее, в лоне математики, как чистой, так и прикладной, а эконометрика, статистика и другие науки применяют выработанные математиками методы.

Пройдя этап параметризации, модель получает «лицо» – коэффициенты при независимых переменных. Однако еще рано применять данную модель на практике или пытаться интерпретировать полученные результаты, так как она должна пройти *этап верификации и валидации*.

По своей природе модель всегда является упрощением реальной действительности, т. е. она всегда менее детализирована, чем представляемая ей система. С одной стороны, такое упрощение и допущения, которые мы делаем для его достижения, устраняют ненужные детали и позволяют нам сосредоточиться на элементах внутри системы, которые нам наиболее важны с точки зрения проводимого исследования.

С другой стороны, этот процесс упрощения вводит в нашу модель некоторую неточность, погрешность. Определенная степень неточности может быть допустимой и даже необходимой и желательной для того, чтобы сделать модельное решение более эффективным. Очевидно, что при построении любой модели неизбежны некоторые предположения и допущения о системе. Однако, сделав такие предположения, мы должны понимать, что придется приложить усилия, чтобы ответить на вопрос о качестве нашей модели.

Есть два шага для оценки того, насколько хорошо модель отражает свойства системы. Мы должны выяснить, правильно ли модель реализует допущения (верификация модели) и были ли сделанные допущения разумными с точки зрения реальной системы (валидация модели).

Важно помнить, что валидация не подразумевает верификацию, а верификация не предполагает валидацию. Однако на практике эти проверки часто смешиваются, особенно когда данные измерений доступны для моделируемой системы. Так, если при сравнении реальных данных и предсказаний модели мы имеем приемлемое расхождение, т. е. результаты, полученные из модели, близки к результатам, полученным в реальности, то предполагается, что построенная модель является и верифицированной, и валидной.

Рассмотрим подробнее процесс валидации. Как уже упоминалось ранее, валидация – это выяснение того, являются ли численные результаты, количественно определяющие отношения между переменными модели (т. е. коэффициенты при переменных), приемлемыми для описания процессов, которые модель направлена отражать. Валидация может включать в себя ряд этапов.

Самым первым шагом на этапе валидации модели может служить оценка коэффициента детерминации. Коэффициент детерминации, обозначаемый R^2 , представляет собой число, которое указывает долю дисперсии в зависимой переменной, которая предсказуема по объясняющей переменной [17].

Этот показатель используется в контексте статистических моделей, основной целью которых является либо прогнозирование будущих результатов, либо проверка гипотез на основе другой соответствующей информации. Он показывает, насколько хорошо наблюдаемые результаты воспроизводятся моделью, исходя из доли полной вариации результатов, объясняемой моделью [18].

Важно понимать, что коэффициент детерминации не может дать ответы на следующие вопросы:

- являются ли независимые переменные причиной изменений зависимой переменной?;
- присутствует ли смещение в оценках коэффициентов модели?;
- была ли использована правильная форма регрессии?;
- являются ли выбранные объясняющие переменные наиболее подходящими?;
- присутствует ли коллинеарность в данных по объясняющим переменным?;
- можно ли улучшить модель путем использования преобразованных независимых переменных?;
- достаточно ли данных, чтобы сделать уверенный вывод?

Кроме того, коэффициент детерминации, близкий к единице, не гарантирует, что модель хорошо подходит для данных, поскольку, как показано в квартете Энскомба [19], высокий коэффициент детер-

минации может возникать в случае неправильной спецификации функциональной формы отношения или при наличии выбросов, которые искажают истинные отношения.

Следующий важный метод валидации модели состоит в анализе ее остатков. Остатки, или отклонения, модели представляют собой различия между расчетным значением показателя, полученным с помощью модели, и реальным его значением, наблюдаемым исторически.

Так, если модель соответствует данным, остатки будут максимально приближаться к величине случайной ошибки. Это означает, что зависимость между объясняющими переменными и зависимой переменной действительно имеет место со статистической точки зрения. Соответственно, если остатки ведут себя беспорядочно, это говорит о том, что модель корректно отражает происходящие процессы. С другой стороны, если в остатках наблюдается присутствие неслучайной структуры, это явный признак того, что модель требует коррекции.

Далее рассмотрим проблемы верификации модели, а конкретнее мультиколлинеарность, автокорреляцию и гетероскедастичность. Эти состояния служат яркими проявлениями нарушения предпосылок теоремы Гаусса – Маркова и негативно влияют на качество модели в целом.

Мультиколлинеарность – явление, в котором две или более предсказывающие переменные в модели множественной регрессии сильно коррелированы. Это означает, что одну из них можно линейно предсказать через другие с высокой степенью точности. В такой ситуации оценки коэффициентов множественной регрессии могут изменяться хаотично в ответ на небольшие изменения в структуре модели или данных. Мультиколлинеарность не ухудшает прогнозные свойства или надежность модели в целом, по крайней мере в пределах набора данных выборки. Она влияет только на вычисления, касающиеся отдельных экзогенных переменных. Иначе говоря, модель множественной регрессии с коррелированными переменными показывает, насколько хорошо весь набор независимых переменных предсказывает зависимую переменную, но она может не давать достоверных результатов ни об отдельной предсказывающей переменной, ни о том, какие из этих переменных являются избыточными в модели.

Одна из особенностей мультиколлинеарности состоит в том, что стандартные ошибки затронутых коэффициентов имеют тенденцию быть завышенными. В этом случае проверка гипотезы о том, что коэффициент равен нулю, может привести к ложному выводу об отсутствии влияния переменной или же ошибки второго рода.

Перейдем к проблеме автокорреляции. Автокорреляция, также известная как последовательная корреляция, представляет собой корреляцию значений данных с отложенной копией себя как функции задержки. Неформально автокорреляцию можно определить как сходство между наблюдениями, приводящее к построению функции временной задержки между ними. Непосредственно в статистике автокорреляция случайного процесса – это корреляция Пирсона между значениями процесса в разное время (как функция от времени).

Автокорреляция остатков нарушает обычное предположение о наименьших квадратах, согласно которому остатки некоррелированы, а это означает, что теорема Гаусса – Маркова неприменима и оценки больше не являются несмещенными. Несмотря на то что автокорреляция не смещает оценки коэффициентов, стандартные ошибки этих коэффициентов, как правило, занижаются (а ошибки t -статистики, соответственно, завышаются), особенно при положительной автокорреляции при малых значениях лагов.

Традиционным критерием наличия автокорреляции первого порядка является статистика Дарбина – Ватсона или, если объясняющие переменные включают в себя запаздывающую зависимую переменную, h -статистика Дарбина. Более гибким тестом, охватывающим автокорреляцию более высоких порядков и применимым независимо от того, включают ли регрессоры лаги зависимой переменной, является тест Бреуша – Годфри. Он включает вспомогательную регрессию, в которой остатки, полученные из оценки модели, регрессируют на исходные регрессоры и k лагов остатков, где k – порядок тестирования.

С другой стороны, в регрессионном анализе с использованием данных временных рядов автокорреляция в интересующей переменной типично моделируется либо авторегрессионной моделью (*autoregressive model*, AR), моделью скользящего среднего (*moving average model*, MA), их комбинацией в виде модели авторегрессионной скользящей средней (*autoregressive moving average model*, ARMA), либо расширением последней, называемой моделью авторегрессии с интегрированной скользящей средней (*autoregressive integrated moving average model*, ARIMA). При использовании нескольких взаимосвязанных рядов данных применяется векторная авторегрессия (*vector autoregression*, VAR) или ее расширения.

При необходимости коррекции автокорреляции можно преобразовать модель или ее переменные одним из следующих способов:

- перейти от модели статической регрессии к модели динамической регрессии путем введения лагов по зависимой или независимым переменным;

- перейти от переменных в абсолютных значениях (уровнях) к приростам, темпам роста или индексам;
- снизить степень коррелированности переменных, перейдя к их логарифмам либо очистив данные от присутствия в них тренда.

Другим решением проблемы автокорреляции является применение авторегрессионной схемы AR. Она является моделью временного ряда, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда.

Обратимся к проблеме гетероскедастичности. В статистике группа случайных величин является гетероскедастичной, если в ней есть подгруппы, которые имеют отличную от других изменчивость. Здесь изменчивость может быть количественно определена дисперсией или любым другим аналогичным показателем. Таким образом, гетероскедастичность – это отсутствие гомоскедастичности. Последнюю мы определим как однородность изменчивости значений наблюдений, которая выражается в гомогенности дисперсии случайной ошибки регрессионной модели.

Существование гетероскедастичности представляет собой серьезную проблему при применении регрессионного анализа, в том числе анализа дисперсии, поскольку это может сделать недействительными статистические тесты значимости, которые предполагают, что ошибки моделирования некоррелированы и однородны, и поэтому их дисперсии не меняются. Например, в то время как обычная оценка методом наименьших квадратов по-прежнему остается несмещенной при наличии гетероскедастичности, она неэффективна, поскольку истинная дисперсия и ковариация в этом случае недооцениваются [20, р. 238–243]. Аналогичным образом при тестировании различий между подгруппами с использованием локального теста некоторые стандартные тесты предполагают, что отклонения в группах равны. Кроме того, присутствие гетероскедастичности упоминается некоторыми авторами как ошибочная спецификация второго порядка [21, р. 66–110].

Гетероскедастичность не вызывает смещения оценок обычного коэффициента наименьших квадратов, хотя может привести к тому, что оценки дисперсий (и, следовательно, стандартные ошибки) коэффициентов будут смещены. Таким образом, регрессионный анализ с использованием данных с присутствием гетероскедастичности будет по-прежнему обеспечивать несмещенную оценку взаимосвязи между независимой и зависимой переменными, но стандартные ошибки и, соответственно, выводы, полученные в результате анализа данных, не могут быть достоверными. Смещенные стандартные ошибки приводят к смещенному выводу, поэтому результаты проверок гипотез, возможно, окажутся неверны. Например, если метод наименьших квадратов выполняется на гетероскедастическом наборе данных, что приводит к необъективной оценке стандартной ошибки, исследователь может не отклонить нулевую гипотезу на данном уровне значимости, когда она фактически нехарактерна для истинной совокупности (совершая таким образом ошибку второго рода). Гетероскедастичность также является важной практической проблемой, возникающей в задачах ANOVA (*analysis of variance*) [22, р. 625].

Стоит, однако, отметить, что наличие гетероскедастичности не критично для определенных случаев. Некоторые авторы утверждают: «Неравномерную дисперсию остатков стоит исправлять только тогда, когда проблема серьезная»² [23, р. 306]. Другие настаивают: «Гетероскедастичность никогда не была причиной того, чтобы отбросить хорошую модель» [24, р. 8].

Тем не менее для любой нелинейной модели (например, моделей Logit и Probit) гетероскедастичность имеет более серьезные последствия: оценки параметров при ее наличии будут смещенными (в неизвестном направлении), а также несогласованными [25]. В связи с этим стоит указать методы анализа и коррекции гетероскедастичности. Таковыми являются тест Левене, тест Голдфелда – Квандта, тест Парка, тест Глейзера, тест Бреуша – Пагана – Годфри, тест Вайта и др.

На рассматриваемом этапе проводится тест Рамсея (RESET), который является общим тестом спецификации для модели линейной регрессии. Данный тест проверяет, помогают ли нелинейные комбинации установленных значений пояснить переменную ответа. Идея, лежащая в его основе, заключается в том, что, если нелинейные комбинации объясняющих переменных обладают какой-либо силой при объяснении зависимой переменной, модель неправильно указывается в том смысле, что процесс генерирования данных может быть лучше аппроксимирован полиномиальным или другими нелинейными функциональными формами [26, с. 350–371].

Наконец, убедившись в приемлемом качестве модели, ее можно применять для конкретных целей, т. е. перейти непосредственно к **этапу применения модели**. Таким образом, мы убедились, что построение эконометрической модели может занимать продолжительное время, особенно в условиях отсутствия специального программного обеспечения. Однако данную последовательность можно прописать в качестве алгоритма для работы программы, которая с некоторыми допущениями будет строить модели анализа и прогноза рисков банковского сектора Республики Беларусь.

²Здесь и далее перевод наш. – В. С., Е. М.

Алгоритм прогнозирования целевых показателей в зависимости от прогнозных значений уровней банковских рисков

Алгоритм составлен в соответствии с обоснованными выше этапами построения эконометрической модели.

В качестве независимых переменных рассматриваются интегральные показатели, отражающие уровни кредитного, валютного, процентного рисков, а также риска ликвидности. Эти показатели представляют собой временные ряды и по своей сути являются индексами, отражающими уровень принятия одноименных рисков конкретными банками и банковским сектором в целом. Динамика интегральных показателей для различных банков будет иметь различные основания, связанные со стратегиями этих банков. Сравнение банков по данным показателям не говорит о том, что какие-то из них лучше или хуже других, например, по уровню кредитного риска.

В качестве целевых показателей (зависимых переменных) выступают показатели рентабельности собственного капитала и достаточности нормативного капитала. Они также представляют собой временные ряды.

Значения названных независимых и зависимых переменных за n прошлых наблюдений алгоритм получает из внутренних баз данных Национального банка Республики Беларусь. По умолчанию период соответствует одному месяцу, а рассматриваемое количество прошлых наблюдений равно 60. Продолжительность периода и количество наблюдений являются параметрами алгоритма, которые могут быть изменены пользователем.

Также пользователь имеет возможность выбирать для прогнозирования другие зависимые и независимые переменные, если по ним доступны данные за достаточное число прошлых периодов.

В дополнение к названным выше параметрам модели пользователь также указывает:

- предпочтительную глубину лагов анализируемых показателей l (по умолчанию $l = 6$);
- ширину доверительного интервала, приемлемую при проверке гипотез;
- дальность прогноза (по умолчанию – три периода вперед).

Шаг 1. Как уже отмечалось, все независимые и зависимые переменные представляют собой временные ряды. Они тестируются на стационарность с применением расширенного теста Дики – Фуллера. В случае необходимости временной ряд может быть скорректирован для получения стационарного ряда (например, алгоритм может перейти от значений ряда в уровнях к значениям в разностях).

Шаг 2. С учетом параметра l (глубина лага) значение целевого показателя в период t рассматривается как линейная регрессия вида

$$TV_t = c + \beta_{ICR} \cdot ICR_{it} + \beta_{ILR} \cdot ILR_{it} + \beta_{IER} \cdot IER_{it} + \beta_{PIR} \cdot PIR_{it} + \varepsilon, \quad (1)$$

где TV_t – значение целевого показателя в момент времени t ; ICR_{it} , ILR_{it} , IER_{it} , PIR_{it} – векторы интегральных показателей кредитного риска, риска ликвидности, валютного и процентного рисков соответственно; β_{ICR} , β_{ILR} , β_{IER} , β_{PIR} – векторы коэффициентов при соответствующих интегральных показателях риска; c – масштабирующий коэффициент; ε – остатки модели.

Каждый из векторов ICR_{it} , ILR_{it} , IER_{it} , PIR_{it} включает значения соответствующего показателя в моменты времени $t, t-1, \dots, t-l$, например:

$$ICR_{it} = \begin{bmatrix} ICR_{it} \\ ICR_{it-1} \\ \dots \\ ICR_{it-l} \end{bmatrix}.$$

Вектор интегрального показателя имеет размерность $(l+1) \cdot 1$. Вектор коэффициентов имеет размерность $1 \cdot (l+1)$. Таким образом, количество независимых переменных в линейной регрессии, выраженной формулой (1), составляет $(l+1) \cdot 4$.

Далее тестом Грейнджера проверяется взаимосвязь независимых и зависимой переменных. После установления наличия причинно-следственных связей по Грейнджеру между переменными алгоритм исключает из линейной регрессии независимые переменные, которые несущественно влияют на целевые показатели.

Затем зависимость последовательно тестируется на наличие мультиколлинеарности. При наличии последней она корректируется в том числе путем удаления мультиколлинеарной переменной из уравнения регрессии и вновь тестируется вплоть до получения приемлемой линейной регрессии.

Этот шаг завершается определением коэффициентов линейной регрессии путем применения метода наименьших квадратов. Таким образом, уравнение линейной регрессии для прогнозирования целевых

показателей в зависимости значений уровней банковских рисков получено. Далее необходимо спрогнозировать будущие значения уровней банковских рисков.

Шаг 3. Массив данных предыдущих периодов для каждого из рассматриваемых интегральных показателей рисков разделяется на две подгруппы, одна из которых при применении машинного обучения используется собственно для обучения, а другая – для контроля. Далее для ряда каждого из интегральных показателей строится несколько прогнозов, основанных на различных методах, при этом используются данные лишь из группы обучения. В целях построения прогнозов для каждого интегрального показателя применяются следующие подходы:

- методы декомпозиции;
- экспоненциальное сглаживание;
- *theta*-метод (метод простого экспоненциального сглаживания со смещением);
- TBATS-метод;
- построение ARIMA-моделей;
- прогнозирование на основе нейронных сетей.

Качество полученных прогнозов оценивается путем сравнения с данными контрольной группы для соответствующих показателей. В качестве оценки ошибки используется критерий MAPE (*mean absolute percentage error*) – средняя абсолютная ошибка в процентах. Для каждого интегрального показателя риска определяется подход, который позволяет получить минимальную ошибку прогноза.

Шаг 4. Для каждого интегрального показателя риска строится прогноз его будущих значений путем применения наилучшего подхода, определенного на шаге 3. При этом используется весь ряд данных (и группа обучения, и контрольная группа). Прогноз строится с учетом заданного пользователем параметра дальности.

Шаг 5. Полученные на шаге 4 прогнозные значения независимых переменных подставляются в уравнение линейной регрессии, и производится расчет прогнозных значений целевых показателей.

Алгоритм работает быстро, так как большая часть подготовительных операций и вычислений также автоматизирована.

Полученные в результате работы алгоритма прогнозные значения целевых показателей могут быть применены, например, для сравнения их с текущими и нормативными значениями. Это позволит предвидеть некоторые проблемы, с которыми могут столкнуться банки, и сформировать рекомендации по их предотвращению.

Заключение

Описанный в данной статье подход к прогнозированию уровня рисков в банковском секторе и влияния рисков на целевые показатели развития банков путем применения эконометрического моделирования и современных алгоритмов машинного обучения может рассматриваться как дополнительный метод анализа состояния указанного сектора. Применяя предложенный в работе алгоритм ко всему банковскому сектору и отдельным банкам, можно получить необходимые сигналы о вероятных кризисных явлениях в банковской сфере и нивелировать негативное влияние надвигающихся событий либо применить его для других определенных пользователем алгоритма исследований.

Таким образом, Национальный банк Республики Беларусь может расширить применяемые на практике сложные механизмы мониторинга и прогноза состояния банковского сектора, что позволит ему подготовить банки к вероятным кризисным явлениям до их непосредственного наступления.

Библиографические ссылки

1. Beck U. Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt : Suhrkamp, 1986.
2. Luhmann N. Soziologie des Risikos. Berlin : Walter de Gruyter, 2003.
3. Koch P. Versicherungsgeschichte in Stichworten // Schriftenreihe des Vereins zur Förderung der Versicherungswissenschaft in München e. v. München, 1988. Vol. 32. S. 1–16.
4. Ritter G. A. Der Sozialstaat: Entstehung und Entwicklung im internationalen Vergleich. München : De Gruyter Oldenbourg, 1991.
5. Crockford N. An Introduction to Risk Management. 2nd ed. Cambridge : Woodhead-Faulkner, 1986.
6. Pabst R. Theorie und Methodenentwicklung bei der Versicherung technischer Risiken am Beispiel der Maschinenversicherung in Deutschland : diss. ... doktors der philos. München, 2011.
7. Первозванский А. А., Первозванская Т. Н. Финансовый рынок: расчет и риск. М. : Инфра-М, 1994.
8. Галова А. Г. Инфляционные риски в Республике Беларусь // Банк. весн. 2013. № 19. С. 8–12.
9. Allen M. P. Understanding Regression Analysis. New York : Springer, 1997.
10. Asteriou D., Hall S. G. Applied Econometrics. London : Palgrave MacMillan, 2011.
11. Timeline of Nobel Prize Winners. Economics [Electronic resource]. URL: <http://www.nobel-winners.com/Economics/> (date of access: 15.04.2017).

12. Granger C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods // *Econometrica*. 1969. Vol. 37 (3). P. 424–438.
13. Diebold F. X. Elements of Forecasting. 2nd ed. Cincinnati : South Western, 2001.
14. Berzuini C. Causal Inference in Time Series Analysis // *Causality: statistical perspectives and applications* / ed. by C. Berzuini. [et al.]. 3rd ed. Hoboken : John Wiley & Sons, 2012.
15. Kannike K. Notes on Feynman parametrisation and the Dirac delta function [Electronic resource]. URL: /http://kodu.ut.ee/~k-kannike/english/science/physics/notes/feynman_param.pdf (date of access: 22.04.2017).
16. Shapiro J. A. Schwinger trick and Feynman Parameters [Electronic resource] // Rutgers. URL: https://www.physics.rutgers.edu/grad/615/lects/schwingertrick.pdf (date of access: 22.04.2017).
17. Statistics and Probability Dictionary [Electronic resource]. URL: http://stattrek.com/statistics/dictionary.aspx?definition=coefficient_of_determination (date of access: 22.04.2017).
18. Glantz S. A., Slinker B. K. Primer of Applied Regression and Analysis of Variance. 2nd ed. New York : McGraw-Hill Educ., 2000.
19. Anscombe F. J. Graphs in Statistical Analysis // *Am. Stat.* 1973. Vol. 27. P. 17–21.
20. Goldberger A. S. Econometric Theory. New York : John Wiley & Sons, 1964.
21. Testing Structural Equation Models / ed. by K. A. Bollen [et al.]. London : Sage, 1993.
22. Jinadasa G., Weerahandi S. Size performance of some tests in one-way anova // *Commun. Stat. – Simul. Comput.* 1998. Vol. 27 (3). P. 625–640.
23. Fox J. Applied Regression Analysis, Linear Models and Related Methods. California : Sage, 1997.
24. Mankiw N. G. A Quick Refresher Course in Macroeconomics // *J. Econ. Lit.* 1990. Vol. 28 (4). P. 1645–1660.
25. Giles D. Robust Standard Errors for Nonlinear Models [Electronic resource] // *Econometrics Beat: Dave Giles' Blog*. URL: http://davegiles.blogspot.com.by/2013/05/robust-standard-errors-for-nonlinear.html (date of access: 25.04.2017).
26. Ramsey J. B. Tests for Specification Errors in Classical Linear Least Squares Regression Analysis // *J. Royal Stat. Soc. Ser. B*. 1969. Vol. 31 (2). P. 350–371.

References

1. Beck U. Risikogesellschaft. Auf dem Weg in eine andere Moderne. Frankfurt : Suhrkamp, 1986 (in Ger.).
2. Luhmann N. Soziologie des Risikos. Berlin : Walter de Gruyter, 2003 (in Ger.).
3. Koch P. Versicherungsgeschichte in Stichworten. In: *Schriftenreihe des Vereins zur Förderung der Versicherungswissenschaft in München e. v.* München, 1988. Vol. 32. S. 1–16 (in Ger.).
4. Ritter G. A. Der Sozialstaat: Entstehung und Entwicklung im internationalen Vergleich. München : De Gruyter Oldenbourg, 1991 (in Ger.).
5. Crockford N. An Introduction to Risk Management. 2nd ed. Cambridge : Woodhead-Faulkner, 1986.
6. Pabst R. Theorie und Methodenentwicklung bei der Versicherung technischer Risiken am Beispiel der Maschinenversicherung in Deutschland : diss. ... doktors der philos. München, 2011.
7. Pervozvanskiy A. A., Pervozvanskaya T. N. Finansovyi ryok: raschet i risk [Financial market: calculation and risk]. Moscow : Infra-M, 1994 (in Russ.).
8. Galova A. G. [Inflation risks in the Republic of Belarus]. *Bankavski vesnik* [Bank Bulletin Magazine]. 2013. No. 19. P. 8–12 (in Russ.).
9. Allen M. P. Understanding Regression Analysis. New York : Springer, 1997.
10. Asteriou D., Hall S. G. Applied Econometrics. London : Palgrave MacMillan, 2011.
11. Timeline of Nobel Prize Winners. Economics. URL: http://www.nobel-winners.com/Economics/ (date of access: 15.04.2017).
12. Granger C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*. 1969. Vol. 37 (3). P. 424–438.
13. Diebold F. X. Elements of Forecasting. 2nd ed. Cincinnati : South Western, 2001.
14. Berzuini C. Causal Inference in Time Series Analysis. In: Berzuini C., Dawid Ph., Bernardinell L. (eds.). *Causality: statistical perspectives and applications*. 3rd ed. Hoboken : John Wiley & Sons, 2012.
15. Kannike K. Notes on Feynman parametrisation and the Dirac delta function. URL: /http://kodu.ut.ee/~kkannike/english/science/physics/notes/feynman_param.pdf (date of access: 22.04.2017).
16. Shapiro J. A. Schwinger trick and Feynman Parameters. In: *Rutgers*. URL: https://www.physics.rutgers.edu/grad/615/lects/schwingertrick.pdf (date of access: 22.04.2017).
17. Statistics and Probability Dictionary. URL: http://stattrek.com/statistics/dictionary.aspx?definition=coefficient_of_determination (date of access: 22.04.2017).
18. Glantz S. A., Slinker B. K. Primer of Applied Regression and Analysis of Variance. 2nd ed. New York : McGraw-Hill Educ., 2000.
19. Anscombe F. J. Graphs in Statistical Analysis. *Am. Stat.* 1973. Vol. 27. P. 17–21.
20. Goldberger A. S. Econometric Theory. New York : John Wiley & Sons, 1964.
21. Bollen K. A., Long J. S., Trivedi P. K., et al. (eds.). Testing Structural Equation Models. London : Sage, 1993.
22. Jinadasa G., Weerahandi S. Size performance of some tests in one-way anova. *Commun. Stat. – Simul. Comput.* 1998. Vol. 27 (3). P. 625–640.
23. Fox J. Applied Regression Analysis, Linear Models and Related Methods. California : Sage, 1997.
24. Mankiw N. G. A Quick Refresher Course in Macroeconomics. *J. Econ. Lit.* 1990. Vol. 28 (4). P. 1645–1660.
25. Giles D. Robust Standard Errors for Nonlinear Models. *Econometrics Beat: Dave Giles' Blog*. URL: http://davegiles.blogspot.com.by/2013/05/robust-standard-errors-for-nonlinear.html (date of access: 25.04.2017).
26. Ramsey J. B. Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *J. Royal Stat. Soc. Ser. B*. 1969. Vol. 31 (2). P. 350–371.

Статья поступила в редколлегию 21.02.2018.
Received by editorial board 21.02.2018.