

**Федеральное государственное бюджетное образовательное
учреждение высшего профессионального образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Каукин А.С., Косарев В.С.

**Моделирование и прогнозирование индексов производства при
помощи искусственных нейронных сетей с учетом
межотраслевых связей и сравнение прогностических качеств
различных архитектур**

Москва 2020

Аннотация

В данной работе проведен анализ возможностей использования сверточных и рекуррентных нейронных сетей для прогнозирования индексов промышленного производства экономики России. Поскольку индексы имеют асимметрию в периодах роста и падения, была выдвинута гипотеза о том, что нелинейные методы позволят улучшить качество прогноза относительно линейных. В практической части исследования были получены свидетельства неотвержения данной гипотезы. Кроме индивидуальных нейросетевых моделей разработана и протестирована система нейросетей, объединяющая процесс прогнозирования во взаимосвязанных отраслях. Важное внимание в работе уделяется методам, которые позволяют интерпретировать результаты, получаемые нейросетевыми моделями. Применяемые подходы позволили провести декомпозицию факторов по степени влияния на вывод сети, а также определить направленность влияния.

This work analyzes the possibilities of using convolutional and recurrent neural networks to predict the indices of industrial production of the Russian economy. We are testing the hypothesis that nonlinear methods will improve the forecast quality relative to linear ones. This assumption is due to the fact that the indices are asymmetric in periods of growth and decline. The hypothesis is not rejected in the practical part of the work. We also developed and tested a neural network system that integrates the forecasting process in related industries. In this work, we pay special attention to methods that allow us to interpret the results of neural network models. We decompose the factors according to the degree of influence on the conclusion and determine the direction of the influence.

Каукин Андрей Сергеевич - к.э.н., зав. лабораторией системного анализа отраслевых рынков ИОРИ РАНХиГС при Президенте РФ

Косарев Владимир Сергеевич – младший научный сотрудник лаборатории системного анализа отраслевых рынков ИОРИ РАНХиГС при Президенте РФ

Данная работа подготовлена на основе материалов научно-исследовательской работы, выполненной в соответствии с Государственным заданием РАНХиГС при Президенте Российской Федерации на 2020 год.

СОДЕРЖАНИЕ

| | |
|---|-----------|
| ВВЕДЕНИЕ..... | 4 |
| 1 Систематизация теоретических и эмпирических подходов к моделированию временных рядов при помощи искусственных нейронных сетей..... | 6 |
| 1.1 Полносвязные нейронные сети..... | 7 |
| 1.2 Рекуррентные нейронные сети..... | 14 |
| 1.3 Сверточные нейронные сети..... | 18 |
| 2 Разработка модели прогнозирования индексов производства отечественной экономики с учетом межотраслевых связей..... | 22 |
| 2.1 Особенности отраслевой динамики и технологические взаимосвязи..... | 22 |
| 2.2 Прогнозирование нелинейных временных рядов..... | 24 |
| 2.3 Определение объясняющих факторов и содержательные гипотезы..... | 30 |
| 3 База данных эмпирического исследования..... | 41 |
| 3.1 Описание структуры базы данных..... | 41 |
| 3.2 Предобработка данных для нейросетевого моделирования..... | 41 |
| 4 Эмпирическое моделирование, верификация полученных результатов..... | 47 |
| 4.1 Проверка гипотез исследования, обучение и выбор оптимальной модели..... | 48 |
| ЗАКЛЮЧЕНИЕ..... | 69 |
| СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ..... | 72 |

ВВЕДЕНИЕ

В данной работе изучается возможность расширения методов моделирования и прогнозирования временных рядов за счет нейросетевых алгоритмов. Основной вопрос, на который предполагается ответить в работе состоит в том, являются ли нейросетевые методы полезными при построении макроэкономических прогнозов.

В практической части исследования предпринята попытка построения сетевой модели, учитывающей отраслевые связи для прогноза основных индексов производства по структуре ОКВЭД. Модель представляет собой комбинацию независимых отраслевых моделей прогнозирования. Необходимость разработки предлагаемой модели обусловлена текущей экономической обстановкой, связанной, в свою очередь, с эпидемиологической – распространением COVID-19. Массовое нарушение производственных цепочек, происходящее в результате вынужденного закрытия части предприятий и невозможности полноценной работы логистики, приводят к кризисным последствиям в отраслях, которые не имели на то экономических предпосылок. Предлагаемая нейросетевая модель способна принимать во внимание подобные риски, поскольку прогноз производства отрасли строится не только с учетом факторов спроса и предложения целевой отрасли, но и с учетом аналогичных факторов для связанной отрасли. При этом прогноз производства целевой отрасли в большей степени опирается на факторы, наследуемые из структурной модели спроса-предложения целевой отрасли, а факторы производства связанной отрасли учитываются опосредованно через прогноз второстепенной модели. Прогноз индексов производства важен для экономических агентов, сталкивающийся с издержками при корректировке определенных макропоказателей и максимизирующий ожидаемое значение дисконтированной будущей прибыли. К примеру, предприниматель, прогнозирующий продажи продукции на основе прогноза индекса промышленного производства, сочтет полезным знать период времени, в течение которого индекс изменит свою траекторию. Кроме того, прогнозы являются ключевыми исходными данными для принятия решений государственными органами. Например, при оценке устойчивости бюджета крайне важно иметь качественные прогнозы выпуска в отраслевом разрезе.

Предлагаемая работа основана на результатах таких исследователей, как Бернс и Митчел, Дж. Блат, С. Нефтичи, Дж. Хамилтон [1, 2, 3, 4]. В этих работах существуют убедительные эмпирические данные, свидетельствующие об асимметричном поведении макроэкономических показателей в периоды отрицательной и положительной динамики. Учет этих явлений имеет основополагающее значение для моделирования временного ряда. Традиционные модели авторегрессии, такие как VAR и ARIMA, не в состоянии уловить нелинейности. Пороговые модели с эндогенным определением порога на основе

марковских цепей зачастую демонстрируют экономически неинтерпретируемые результаты. Нейронные сети, не требуя априорного знания вида функциональной связи, способны учитывать нелинейные связи. Однако множество архитектур не могут быть использованы, поскольку требуют «больших данных», которые отсутствуют в макроэкономической статистике. Хотя различные архитектуры нейронных сетей использовались для прогнозирования макропоказателей: валютного курса, ВВП, инфляции, безработицы [5, 6, 7, 8], отечественная статистика требует особых подходов, учитывающих ее сравнительно короткую историю наблюдений.

Сверточные сети в отличие от нейронных сетей прямого распространения обладают двумя важными достоинствами: для получения стабильного результата прогнозирования не требуется большой объем выборки данных и к тому же учитывают пространственно-временной порядок данных. В основе, предлагаемой сверточной нейронной сети находится архитектура «WaveNet» [9], разработанная для анализа и прогнозирования звукового сигнала. Данная архитектура уже была адаптирована для анализа временных рядов и применялась при прогнозировании фондового рынка [10], а так же поворотных точек индексов производства в НИР «Идентификация и прогнозирование фаз российского экономического цикла с учетом отраслевой структуры экономики при помощи искусственных нейронных сетей» 2019 года. В практической части работы предполагается использовать имеющиеся архитектурные подходы с учетом их адаптации для прогнозирования значений уровней индексов. При этом одной из ключевых задач исследования является построение комбинированной модели на основе индивидуальных нейросетевых прогнозных отраслевых моделей, в соответствии с имеющимися производственными взаимосвязями.

Статья имеет следующую структуру. В первой главе исследования приведен обзор литературы в области применения нейросетевых методов к задачам прогнозирования временных рядов. Во второй главе исследования изложена концепция собственной прогнозной модели. В третьей главе исследования приведено описание имеющейся базы данных. В четвертой главе исследования приведены результаты построения нейросетевых моделей с использованием различных архитектур и их сравнительные прогностические характеристики.

1 Систематизация теоретических и эмпирических подходов к моделированию временных рядов при помощи искусственных нейронных сетей

Искусственные нейронные сети (ИНС) являются одной из наиболее точных и широко используемых моделей прогнозирования, которые нашли широкое применение в прогнозировании социальных и экономических показателей. Можно выделить ряд отличительных особенностей искусственных нейронных сетей, которые делают их привлекательными для задачи прогнозирования. Во-первых, в отличие от традиционных методов, основанных на структурных моделях, искусственные нейронные сети являются методами, самоадаптируемыми посредством данных, и нуждаются в относительно небольшом количестве априорных предположений относительно исследуемой проблемы. Во-вторых, искусственные нейронные сети обладают высокоэффективными обобщающими способностями. На основе обучающей выборки нейронные сети часто могут правильно построить функцию распределения данных, даже если данные содержат зашумленную информацию. В-третьих, ИНС являются универсальными функциональными аппроксиматорами. Теоремы Сибенко и позже ее обобщенная версия Хорника [11] говорят о том, что нейронная сеть может аппроксимировать любую непрерывную функцию с любой требуемой точностью. Наконец, искусственные нейронные сети являются нелинейными. Традиционные подходы к прогнозированию временных рядов, такие как VAR или ARIMA, предполагают, что исследуемые временные ряды генерируются из линейных процессов. Однако подобное предположение может быть некорректным, если механизм генерации является нелинейным.

В экономической практике нейронные сети использовались для прогнозирования валютного курса, ВВП, инфляции, безработицы [5, 6, 7, 8]. В том числе центральные банки Англии [12], Канады [13] и ФРС США [14] применяют их для различных целей. Банк Англии использует нейросеть для прогнозирования баланса финансовых организаций для целей банковского надзора, а также для прогнозирования инфляции на среднесрочный горизонт до двух лет. Кроме того, Банк Англии заявлял о использовании нейросетевых алгоритмов в исследованиях схем финансирования технологических стартапов с целью выявления революционных изменений в финансовой индустрии. Банк Канады использует нейронную сеть для прогнозирования ВВП. ФРС США использует четыре различных архитектуры для прогноза инфляции и безработицы.

При этом нейронные сети – это широкий спектр принципиально различных алгоритмов. Их объединяет, как правило, только общий подход к обучению, но архитектура построения и принципы работы различны. В применении к анализу и

прогнозированию временных рядов представляется логичным рассмотреть три базовые концепции нейронных сетей: полносвязные сети, рекуррентные сети и сверточные сети.

1.1 Полносвязные нейронные сети

Полносвязные нейронные сети прямого распространения (FNN – Fully Connected Neural Network) – это один из самых ранних методов машинного обучения. Их история берет свое начало с описания первого формального нейрона Уолтером Питтсом и Уореном Маккалоком в 1943 году и проведенного в 1956 году Дармутского семинара, ознаменовавшего первую волну научного и общественного интереса к искусственному интеллекту. Однако приближенное к современному использование имеющихся наработок стало возможным только с появлением алгоритма обратного распространения ошибки в 1980 году. Более полную историю можно прочитать в книге [11]. Здесь же стоит заметить, что методы машинного обучения, в том числе и нейронные сети – это новое применение известных математических наработок и методов с относительно усовершенствованными компьютерными технологиями. Глубокая полносвязная нейронная сеть представляет собой, в упрощенном варианте, пул линейных регрессий, связанных между собой одновременно параллельно и последовательно. Отсюда может возникнуть вопрос: почему совокупность регрессий называется машинным обучением? Возможно, и одну регрессию можно считать машинным обучением, если параметры регрессии оценивать не по известной аналитической формуле метода наименьших квадратов, а минимизировать функцию ошибки методом градиентного спуска. По классическому определению, машинное обучение – это «обучение компьютерной программы по мере накопления опыта относительно некоторого класса задач T и целевой функции P , когда качество решения этих задач улучшается с получением нового опыта» [15]. Один нейрон в сети – это линейная регрессия, результат которой пропущен через некоторую нелинейную функцию. Поэтому в данном случае можно говорить в терминах классической эконометрики: имеется некоторая случайная величина Y , математическое ожидание которой это функция от факторов X и параметров b , что и есть нейронная сеть. Пример такой глубокой (двухслойной) полносвязной нейронной сети приведен на рисунке . В данном случае на вход нейронной сети поступает три фактора, линейная комбинация которых с различными весами – это выходы нейронов первого уровня. В свою очередь выходы нейронов первого уровня в виде линейных комбинация являются выходами второго уровня нейронов. Результат нейронной сети – это линейная комбинация результатов, полученных на последнем уровне с некоторыми весами. Веса для линейных комбинаций каждого уровня определяются в результате минимизации функции ошибки посредством специально

разработанного алгоритма, позволяющему эффективно оптимизировать нелинейные, цепочно-вложенные функции, что и представляет собой функция ошибки.

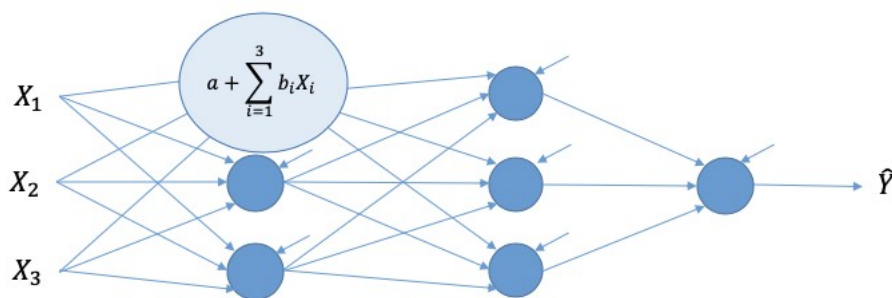


Рисунок 1 – классическая нейронная сеть прямого распространения

Область машинного обучения, по мере оформления в самостоятельную науку, наполняется профессиональными терминами, кратко описывающими громоздкие процедуры, состоящие из давно разработанных математических методов, которые необходимы для обучения нейронных сетей. К примеру, слой нейронной сети – это первый уровень нейронов (регрессий), скрытый слой – второй уровень регрессий, который принимает на вход результаты первого уровня, метод обратного распространения ошибки – это оптимизация функции ошибки, основанная на методе градиентного спуска. Концепция слоя очень важна с вычислительной точки зрения. Вычисления в целом слое можно векторизовать, что необходимо для ускорения расчетов и является необходимым при наличии больших объемов данных. Важно заметить, что в настоящее время существует множество улучшений стандартных методов оптимизации, позволяющих считать быстрее и точнее, но это скорее стоит рассматривать, не как революционные преобразования, а некоторые доработки. Однако в совокупности с высокопроизводительными процессорами, в том числе графическими мощными процессорами, эти технологии позволяют совершать прорывы в разных областях – компьютерном зрении (сегментация, распознавание изображений), обработка и генерация текстов, видео, звука. Возвращаясь к временным рядам, важно заметить, что представление нейронной сети, как пула регрессий – это существенное упрощение, поскольку не принимается во внимание главная особенность нейронных сетей – возможность учета нелинейных связей, когда основная задача нейронной сети – это определять неочевидные, нелинейные зависимости между данными, относительно которых исследователь может не иметь априорных представлений или не имеет возможности их задать в аналитическом виде. Учитывая вышесказанное, представление используемой, в качестве примера, полносвязной глубокой сети изменится следующим образом: линейная комбинация факторов на каждом уровне, перед использованием в

качестве входных данных последующим слоем, проходит через нелинейную функцию активации, в данном случае сигмоид, уравнение (1).

$$f_{act} = \frac{1}{1 + e^{-\left(\sum_{i=1}^3 a_i Out_i + c\right)}} \quad (1)$$

где

f_{act} – сигмоидальная функция;

a_i, c – параметры;

Out_i – выход последнего слоя нейронной сети.

Нелинейная функция активации позволяет переводить данные в новое подпространство, где они линейно разделимы. Рисунок демонстрирует, что каждый нейрон сети, кроме последнего, включает в себя сигмоидальную функцию активации, которая применяется к линейной комбинации входящих векторов. Функциональный вид последнего нейрона определяет какую задачу решает нейронная сеть.

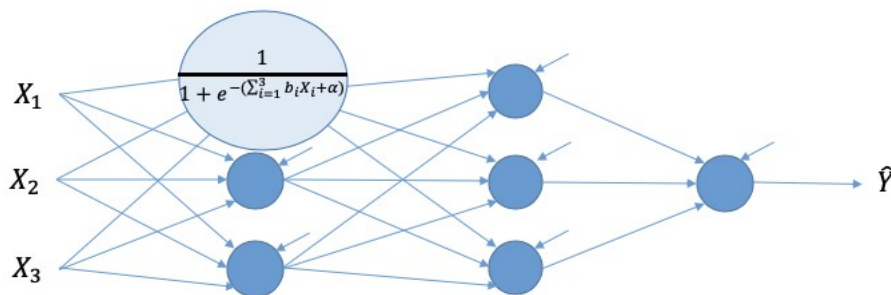


Рисунок 2 – полносвязная сеть прямого распространения с сигмоидальной функцией активации

Любая нейронная сеть решает задачу классификации или регрессии. При работе с изображениями, видео, текстами – это многоклассовая классификация как правило, при работе со звуком – это регрессия. В случае регрессии, последний нейрон имеет линейную функцию активации (уравнение (2)). Аналогично линейной регрессии, функция ошибки представляет собой сумму квадратов отклонения прогноза от фактических данных (уравнение).

$$f_{act} = \sum_{i=1}^3 b_i Out_i + c \quad (2)$$

$$E = \sum (y - \hat{y})^2 \quad (3)$$

где

f_{act} – прогноз нейронной сети;

b_i, c – параметры;

Out_i – выход последнего слоя нейронной сети;

E – функция ошибки;

y, \hat{y} – соответственно фактические значения и прогнозируемые.

Для решения задачи классификации достаточно заменить функцию активации последнего нейрона с линейной на сигмоидальную, в результате чего выход нейрона можно интерпретировать, как вероятностный. Внутренне устройство сети аналогичное - линейные комбинации признаков и нелинейные функции активации (сигмоиды). Соответственно, меняется функция ошибки. Вместо суммы квадратов отклонений возникает потребность минимизировать кросс-энтропию, как в стандартной логистической регрессии, показывающую схожесть двух дискретных распределений (уравнение).

$$E = -\frac{1}{N} \sum_{i=1}^N \ln \hat{y}_i \ln y_i \quad (4)$$

где

E – функция ошибки;

N – количество наблюдений;

w – вектор параметров;

y, \hat{y} – соответственно фактические значения и прогнозируемые.

Таким образом, нейронная сеть с вероятностным выводом в конечном итоге сводится к стандартной линейной логистической регрессии, но тем не менее все преобразования данных внутри сети, происходящие до конечной сигмоиды, позволяют перевести признаки в линейно разделимое пространство. Продемонстрировать сказанное можно на примере общеизвестной задачи бинарной классификации Хог (исключающего или), в которой один класс точек лежит в первой и третьей четвертях, а второй класс – во второй и четвертой. Линейные модели в данном случае будут строить тривиальную разделяющую гиперплоскость. К примеру, стандартная логистическая регрессия (левое изображение рисунка) проведет прямую, относя все что выше ко второму классу, все что ниже к первому. В случае использования трехслойной нейронной сети прямого распространения задача будет решена более корректно (правое изображение рисунка).

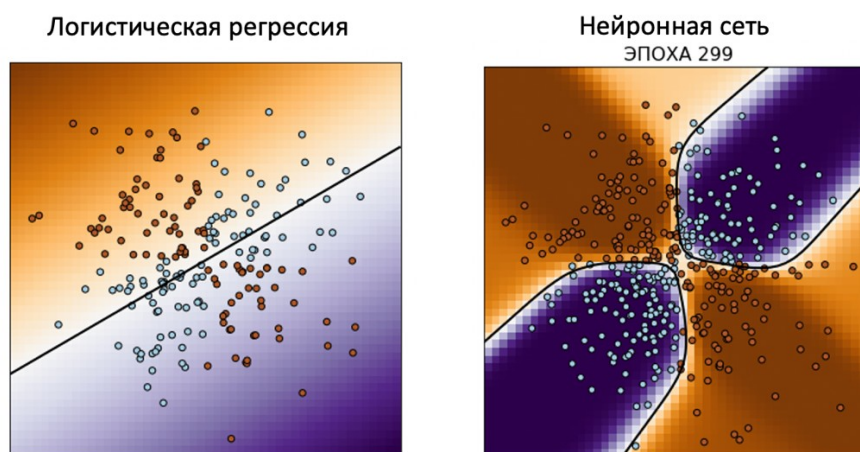


Рисунок 3 – решение задачи «исключающего или» логистической регрессией и полносвязной трехслойной нейронной сетью прямого распространения

Данный пример также демонстрирует необходимость конструирования глубоких сетей. Несмотря на то, что в соответствии с упомянутой выше теоремой Хорника, любую непрерывную функцию можно сколь угодно точно приблизить нейронной сетью с одним слоем, в большинстве практических реализаций нейросетей используются глубокие сети. На практике это позволяет экономить вычислительные мощности, поскольку глубокая сеть с общим количеством нейронов, равным однослойной сети, обучается быстрее и имеет большую обобщающую способность. Теоретическое доказательство такого утверждения было выполнено учеными из массачусетского технологического университета (MIT) [16]. Рольник и Тегмарк доказали полезность глубины, заставляя нейросети выполнять задачу перемножения полиномиальных функций. Это уравнения с переменными, возведёнными в натуральные степени, такими как, $y = x^3 + 1$. Сети тренировались, на примерах уравнений и результатах их перемножения, затем нейросети вычисляли результат умножения уравнений из тестовой выборки, решения которых не были известны заранее. Более глубокие нейросети научились делать это при помощи меньшего общего количества нейронов.

Возвращаясь к примеру «исключающего или», стоит заметить, что на практике встречаются более сложные задачи, однако существуют доказательства [17] того, что двухслойная сеть позволяет отделить произвольный выпуклый многогранник. Трехслойная сеть, в свою очередь, позволяет отделить произвольную многогранную, не обязательно выпуклую, и даже не обязательно связную фигуру.

Попытки применения полносвязных сетей прямого распространения в задачах прогнозирования временных рядов были осуществлены еще в конце XX века. Нейронная сеть, как правило, инкорпорировалась в гибридную модель прогнозирования временных рядов ARIMA - ANN. Использование гибридной модели или объединение нескольких моделей стало обычной практикой для повышения точности прогнозирования с выходом

работы [18] в 1982 году, в которой было продемонстрировано, что комбинация прогнозов из нескольких моделей приводит к повышению эффективности прогнозирования. В статье Клемен [19] предоставлен всеобъемлющий обзор ранних работ в этой области. Основная идея комбинирования моделей в прогнозировании, как указывалось, заключается в использовании уникальных возможностей каждой модели захватывать различных паттерны поведения в данных.

В работе [20] гибридная система состоит из двух этапов. На первом этапе модель ARIMA используется для анализа линейной части ряда. На втором этапе используется модель нейронной сети для моделирования остатков от модели ARIMA. Таким образом, моделирование линейных и нелинейных связей происходит отдельно, после чего прогноз модели ARIMA дополняется за счет спрогнозированных остатков при помощи нейронной сети. При этом отмечается, что при построении моделей ARIMA и ANN требуется субъективная оценка порядка и адекватности модели. В работе отмечается, что более эффективно комбинировать отдельные прогнозы, основанные на различных информационных наборах. В практической части работы моделировался обменный курс британского фунта к доллару США. Используемые данные содержали еженедельные наблюдения с 1980 по 1993 год, или 731 наблюдение. Результаты показали, что для краткосрочного прогнозирования (1 месяц) и нейронная сеть, и гибридные модели превосходят по точности модель случайного блуждания. Для более длительных периодов времени модель ANN дает сопоставимую производительность с моделью ARIMA. Гибридная модель превзошла модели ARIMA и ANN на трех различных временных горизонтах, выбранных в качестве тестовых, согласно используемым показателям погрешности: среднеквадратичной ошибке и среднему абсолютному отклонению.

Аналогичный подход был использован в работе [5]. Практической целью работы являлось также прогнозирование обменного курса. Мотивация применения гибридного подхода состояла в необходимости создания модели, способной превзойти прогноз случайного блуждания. Авторы использовали данные обменных курсов валют: евро/доллар, британский фунт/доллар, и японская йена/доллар. Значения курсов валют были взяты с 4 января 1999 по 30 декабря 2015 и усреднены до дневных (4 242 наблюдения). Идейно построение прогноза в данной работе аналогично описанному выше, но вместо ARIMA использовалась векторная модель коррекции ошибок в случае наличия коинтеграции рядов и векторная авторегрессионная модель в противном случае.

Лучшая линейная модель, согласно критерию Акаике, использовалась в качестве бенчмарка, как и модель случайного блуждания, а также в качестве первичной модели для гибридного подхода. Нейронная сеть прямого распространения использовалась для

моделирования остатков VECM/VAR модели. Устройство гибридной VAR/VECM-ANN модели приведено на рисунке .

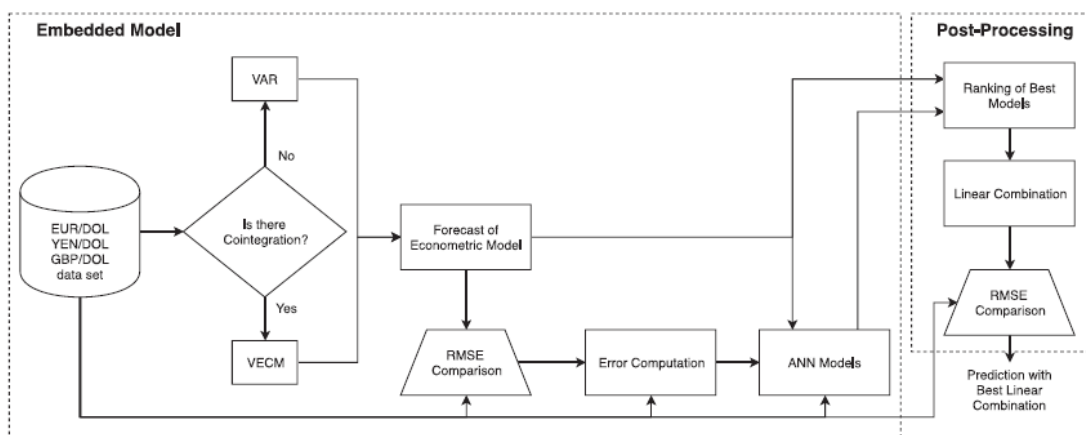


Рисунок 4 – устройство гибридной VAR/VECM-ANN модели из работы [5]

Конечный прогноз строится по описанной выше двухшаговой процедуре: нейронная сеть строит прогноз ошибки линейной модели, который затем добавляется в прогноз VAR/VECM модели. Для нейронной сети, в качестве размерности входных данных, авторы использовали последовательно окна различной длины: 42 дня рабочих дня (два месяца), 63 дня (три месяца), 84 дня (четыре месяца), 252 дня (год), и 378 дней (один год и три месяца). При этом использовалась двухслойная и однослойная сети с различным количеством нейронов на слоях (от 5-ти до 20). В результате комбинированная модель не продемонстрировала устойчиво лучшего результата. Хотя, комбинированная модель с двухслойной сетью и всеми используемыми комбинациями нейронов на слоях оказалась лучше на вневыборочном прогнозе по ошибке отклонения против линейных моделей и нейронной сети в чистом виде.

Недостатками подхода, описанного в работах выше, является переобучение нейронной сети, вероятность чего повышается по мере сокращения выборки и увеличения числа нейронов количества слоев, а также невозможность учесть одновременно пространственную и временную корреляцию в данных. Рассмотрим в качестве примера одну архитектуру нейронной сети из последней работы [5]. Согласно описанию, это двухслойная сеть прямого распространения с окном 42 дня, 5 нейронами на первом слое и 5 нейронами на втором слое. В такой полносвязной сети фрагмент окна каждого фактора является частью входа. Визуализация архитектуры такой сети приведена на рисунке . В модели использовалось три фактора, следовательно входной вектор имеет размерность, равную 126 ($42 \cdot 3 = 126$). Количество обучающих образцов 4 199 ($4\,242 - 42 + 1 = 4199$). При этом на первом слое из 5-ти нейронов должно быть 635 весовых коэффициентов ($126 \cdot 5 + 5 = 635$). Из которых 630 весов сумматора и 5 весов смещения. На втором слое 30 ($5 \cdot 5 + 5 = 30$), и выход нейронной сети, представляющий собой сумматор без функции

активации, поскольку решается задача регрессии, добавляет 10 параметров. Суммарное количество параметров, которое необходимо обучить в процессе минимизации ошибки равняется 675. Таким образом, на один параметр приходится 6.22 наблюдения. При этом рекомендуемое соотношение количества обучаемых параметров к количеству обучающих образцов, не приводящее к переобучению, должно быть больше/равно десяти. Следующей проблемой такого вида архитектуры является невозможность качественного учета пространственной и временной корреляции одновременно. Использование временного окна в 42 дня, предполагает, что нейронная сеть определит нелинейные паттерны поведения на данном временном периоде. При этом нейронная сеть принимает на вход три фактора, как единый вектор, размерности 126, что не дает возможности учесть в явном виде пространственную структуру.

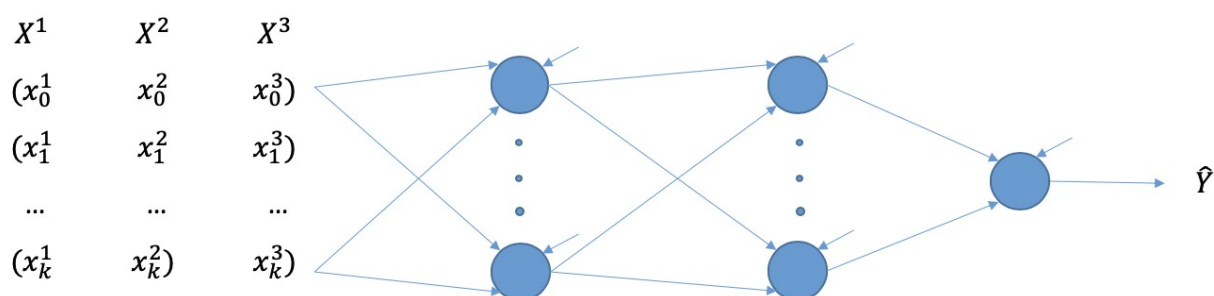


Рисунок 5 – нейронная сеть прямого распространения, учитывающая временную структуру ряда

1.2 Рекуррентные нейронные сети

Рекуррентная сеть – это тип нейронной сети, используемый для временных рядов или последовательных данных, таких как текст и речь. Привлекательность рекуррентных нейронных сетей (RNN) состоит в том, что они потенциально умеют связывать предыдущую информацию с текущей. В отличие от полносвязных нейронных сетей, рассмотренных в предыдущем разделе, RNN используют внутреннюю память для обработки входных данных. Глубокие рекуррентные сети являются предпочтительными из-за их способности учитывать закономерности на длительном временном интервале. Рекуррентные сети используются для анализа и моделирования временных рядов в различных областях (распознавание рукописного ввода, распознавание речи и т. д.). Данный тип сетей хорошо подходит для прогнозирования следующего символа в тексте, используется в приложениях для машинного перевода и последовательной обработки данных [21].

Архитектура модели RNN состоит из разного количества уровней и разных типов блоков на каждом уровне. Основное различие между рекуррентными и полносвязными сетями состоит в том, что каждый блок RNN принимает текущие и предыдущие входные

данные одновременно, обрабатывая входные последовательности одну за другой в каждый момент времени во время работы. Таким образом, выходные данные зависят от предыдущих данных в модели RNN. Рекуррентная нейронная сеть может рассматриваться как несколько копий одной и той же сети, каждая из которых передает данные преемнику.

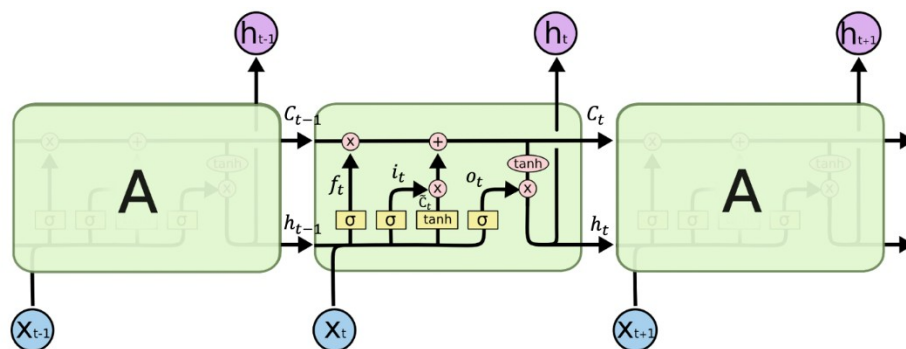
Петля в «неразвернутой» ячейке называется «вектором состояния» и хранит информацию об истории ввода, передавая часть предыдущего состояния ячейки в будущее. Предыдущая и актуальная информация в ячейке взвешивается, после чего сумма пропускается сквозь нелинейную функцию активации, в данном случае гиперболический тангенс. Параметры такой сети обучаются методом обратного распространения ошибки, аналогично нейронной сети прямого распространения, однако из-за проброса градиентов по времени назад возникает ряд сложностей.

При обратном распространении текущие градиенты пробрасываются во времени назад, и если градиент не взрывается, он постепенно затухает. Получается, что влияние текущего слоя не может во времени проброситься слишком далеко назад. Влияние текущего слоя затухает экспоненциально по мере удаления и мешает обычным рекуррентным сетям находить в данных «далекие» зависимости. В результате, память, реализованная подобным образом, получается весьма короткой. Поскольку каждый раз прошлая информация в рекуррентной ячейке смешивается с новой и спустя 5-7 итераций уже полностью перезаписывается. Это очень условное описание данного явления, на самом деле это фундаментальная проблема нейронных сетей, которая называется проблемой затухающих или взрывающихся градиентов. Если матрица весов такова, что заметно увеличивает норму вектора градиента при проходе через один «виртуальный слой» обратного распространения, получится, что при проходе через T слоев эта норма возрастет экспоненциально от T , поскольку вес будет один и тот же во временной петле. Данная проблема была подробно исследована в работе [22].

Необходимость решения данной проблемы породила создание нового типа рекуррентной сети с долговременной памятью, обычно называемой «LSTM» (Long Short Term Memory networks). LSTM и ее вариации наряду с некоторыми гибридными моделями доминируют в области прогнозирования финансовых временных рядов. Такая сеть представляет собой особый тип RNN, способный обучаться долгосрочным зависимостям. Впервые описание ячейки LSTM появились в работе [23], после чего последовало множество дополнений и улучшений первоначальной версии.

Внутри рекуррентной ячейки долговременная память моделируется явным образом. Стоит отметить, что с появлением возможности учитывать долговременные зависимости произошло существенное увеличение количества параметров, которые необходимо

обучить. Сеть LSTM также, как и RNN имеют цепочечную структуру, но повторяющийся модуль имеет более сложное строение. Вместо одного слоя нейронной сети существует четыре, взаимодействующих совершенно особым образом (рисунок).



Примечание – источник [24]

Рисунок 6 – структура ячейки рекуррентной рекуррентной сети с долгосрочной памятью

Вектор верхнего уровня или вектор долгосрочной памяти (C_{t-1}), передающий информацию в реплику ячейки, хранит информацию о предыдущих состояниях. С течением времени не предполагается его существенной мутации. Вектор нижнего уровня или вектор краткосрочной памяти (h_{t-1}) зависит от прошлых значений сети, но может значительно меняться от текущего входа и подает на вход сиюминутную реакцию от текущего набора данных. Первая сигмоида, результатом которой является f_t , отвечает за часть информации из предыдущего состояния, которое нужно «забыть» (уравнение). Поскольку область определения сигмоиды от 0 до 1, функционально данный блок определяет, какую часть от предыдущего значения сохранить для будущего. Следующая сигмоида (уравнение) и гиперболический тангенс (уравнение) совместно определяют какое количество информации от текущего входа передается в вектор долгосрочной памяти. Функция гиперболического тангенса – это классическая рекуррентная ячейка внутри LSTM, а сигмоида определяет, какую долю информации необходимо передать в вектор состояния. Текущее состояние вектора долгосрочной памяти (уравнение), складывающееся из части сохраненной информации от предыдущего шага и части информации от текущего входа, передается в следующую реплику ячейки. Обновление вектора h_t происходит путем перемножения o_t (уравнение), что есть нелинейная комбинация предыдущего состояния вектора краткосрочной памяти (h_{t-1}) и текущего входа (x_t), и гиперболического тангенса от текущего состояния вектора долгосрочной памяти – C_t (уравнение 14). Таким образом, новый h_t в большей степени зависит от текущего значения, но краткосрочная память h_{t-1} и долгосрочная память C_t – оказывают на него некоторое влияние.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (6)$$

$$\check{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (7)$$

$$C_t = f_t * C_{t-1} + i_t * \check{C}_t \quad (8)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = \tanh(o_t * \tanh(C_t)) \quad (10)$$

где

f_t – доля информации, которую нужно «забыть»;

W – матрица весов;

h_{t-1} – краткосрочная память ячейки;

x_t – новые данные;

C_t – вектор долгосрочной памяти;

i_t – доля долгосрочной памяти, передающаяся в следующую реплику ячейки;

o_t – доля краткосрочной памяти, передающаяся в следующую реплику ячейки.

Важно заметить, что в рекурсивном изменении состояния ячейки долгосрочной памяти отсутствует нелинейность. Именно эта особенность позволяет решить проблему исчезающих градиентов. Однако проблема взрывающихся градиентов остается.

Таким образом, ключевой элемент LSTM это состояние ячейки. Она проходит напрямую через цепочку, участвующую лишь в нескольких линейных операциях. Информация может легко течь по ней не подвергаясь преобразованиям. Фильтры контролируют поток информации и могут удалять лишнюю. Они состоят из слоя сигмоидальной нейронной сети и операции умножения. Сигмоида или гиперболический тангенс возвращают число в диапазоне от нуля до единицы, указывая какую долю информации нужно сохранить для будущего. В LSTM три таких фильтра контролируют состояние ячейки. Часть забывается, часть берется из нового входа. Все эти манипуляции делают ячейки очень гибкими. Однако, как было указано выше, данный тип сети требует обучения большого количества параметров. В случае наличия короткой истории наблюдений временного ряда, что является обычной практикой в макроэкономике, неизбежна проблема переобучения. Структура нейросети, которая будет рассмотрена в следующем разделе, во многом сохраняет в себе преимущества рекуррентной сети, но использует на порядок меньше параметров.

1.3 Сверточные нейронные сети

Сверточные нейронные сети (CNN – Convolutional Neural Networks) – это тип глубоких нейронных сетей, состоящий из последовательности сверточных слоев, выходы которых связаны только с локальными областями на входе. Такое устройство реализуется посредством скольжения фильтра или матрицы весов по входным данным с вычислением в каждой точке точечного произведения между ними (т.е. свертки между входным сигналом и фильтром). Эта структура позволяет модели обучать фильтры, которые способны распознавать определенные шаблоны во входных данных. Это наиболее распространенная модель для задач классификации и обработки изображений (классификация изображений, обнаружение объектов, сегментация изображений и т. д.). Преимущество данного типа сети заключается в количестве параметров, которое требуется обучить, по сравнению полносвязными сетями прямого распространения или рекуррентными сетями при решении однотипной задачи. Кроме того, меньшее количество параметров выгодно для оптимизации вычислений и хранения. Идея применения CNN для прогнозирования временных рядов состоит в том, чтобы обучить фильтры, которые способны выделять определенные повторяющиеся шаблоны в данных, и использовать их для прогнозирования будущих значений. Благодаря многоуровневой структуре CNN могут хорошо работать с зашумленными данными, отбрасывая в каждом последующем слое шум и выделяя только значимые шаблоны, имея сходство, таким образом, с нейронными сетями, которые используют вейвлет-преобразованные временные ряды.

Вейвлет-преобразования также непосредственно используются для прогнозирования временных рядов в финансово-экономических задачах. Вейвлет-преобразование позволяет разложить временной ряд на ортогональные компоненты на разных частотах, которые, в свою очередь, локализованы во времени (рисунок). В работе [25] продемонстрирована способность вейвлетов моделировать циклические и частотные свойства цен. Применяя технику вейвлетов, авторы исходили из предположения, что одни факторы влияют на колебания в краткосрочном периоде, другие приводят к изменению тенденции.

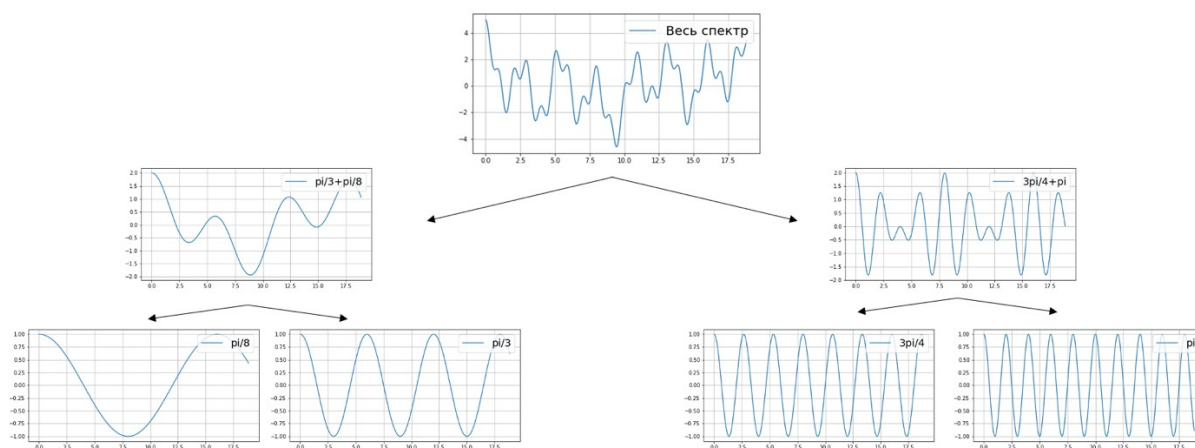


Рисунок 7 – пример спектрального анализа временного ряда

Структура сверточной сети тесно связана с дискретным вейвлет преобразованием. Вейвлет может быть использован для того, чтоб понимать, как имеющаяся функция изменяется от периода к периоду путем сопоставления вейвлет функции с различными участками исходного временного ряда. DWT – это линейная трансформация $x = (x_t)_{t=0}^{N-1}$ с $N=2^J$ с декомпозицией сигнала на высокочастотную и низкочастотную компоненты со сжатием данных с высоко и низко частотным фильтрами. В частности, на каждом уровне j преобразования входной сигнал разлагается на вейвлет $N_j = \frac{N}{2^j}$ и масштабирующие коэффициенты $\langle x, \psi_{j,k} \rangle$ и $\langle x, \phi_{j,k} \rangle$ (также называемый приближением и детализацией) для $k=0, \dots, N_j-1$, путем одновременной свертки входа x фильтрами h и g , заданными как:

$$h(t) = 2^{\frac{-j}{2}} \psi(-2^{-j}t), \quad (11)$$

$$g(t) = 2^{\frac{-j}{2}} \phi(-2^{-j}t), \quad (12)$$

где

$\psi(\cdot)$ – вейвлет;

$\phi(\cdot)$ – функция масштабирования.

На каждом последующем уровне мы применяем преобразование к коэффициентам аппроксимации, таким образом отбрасывая высокочастотные компоненты (детализацию) и заканчивая сглаженной версией входного сигнала. Это очень похоже на структуру CNN, где на каждом последующем уровне мы объединяем выход предыдущего уровня с обучаемым фильтром. В каждом слое фильтр используется для распознавания локальных зависимостей в данных, которые впоследствии объединяются для представления более

глобальных признаков, пока на последнем уровне мы не вычислим интересующий результат.

Исходный ряд будет меньшей размерности, если не используется закичивание. Однако в случае временных рядов его использовать нецелесообразно, поскольку усреднять данные в конце ряда с данными в начала ряда не имеет смысла. Эта технология сохранения размерности полезна только в случае сжатия изображений, где вейвлеты активно используются. Но идея применения вейвлетов для прогнозирования временных рядов состоит в том, чтоб прогнозировать частоты одного уровня, а потом восстанавливать исходный ряд обратным вейвлет преобразованием. Здесь можно привести аналогию с методом, когда прогнозируются отдельно тренд, сезонность, а потом результат складывается. Только в случае вейвлетов в уравнении регрессии и слева и справа будут стоять вейвлеты одного порядка частотности. Поэтому обратное вейвлет преобразование приводит к исходной размерности данных.

В случае же сверточной сети мы используем тот же инструмент для обработки временного ряда – фильтры. Но в случае вейвлетов мы имеем некоторый перечень фильтров с заданными весами и просто знаем, какие фильтры какие диапазоны частот способны улавливать. Так, к примеру, определенные фильтры способны выделить из сигнала, содержащего колебания частотой от 0 до π , колебания в диапазоне от 0 до $\frac{\pi}{8}$ и т.д.

В сверточной же сети фильтры являются обучаемыми, их параметры определяются в результате оптимизации целевой функции таким образом, что в конечном итоге они выделяют только «полезные» диапазоны частот для предсказания целевой переменной, отбрасывая ненужную информацию. Именно по этой причине в литературе, посвященной применению сверточных нейронных сетей к временным рядам, часто встречается утверждение о том, что сверточные сети способны хорошо работать с зашумленными и некачественными данными. В макроэкономике, где ошибки измерения и погрешности при вычислении агрегатных факторов могут быть велики, применение такого способа представляется полезным.

Возвращаясь к моделированию временных рядов посредством сверточных сетей рассмотрим некоторые примеры. Галещук и Мукерджи [26] использовали сверточную сеть для прогнозирования направления изменения курса валют и биржевых индексов. Согласно результатам сверточная сеть продемонстрировала относительно высокую точность классификации по сравнению с выбранными бенчмарками: моделью опорных векторов, и полносвязной нейронной сетью с одним скрытым слоем. В работе [27]

сверточная сеть используется для прогнозирования временных рядов, где пространственно-временные данные потока трафика пассажиров представлены в виде изображений. Модель CNN прогнозирует скорость движения в крупных транспортных сетях. Среди преимуществ сети отмечается ее возможность учитывать пространственные отношения между соседними временными рядами. В работе Миттельмана [28] прогнозируются биржевые индексы с применением сверточной нейронной сети, которая использует одномерные свертки.

Нейронные сети – это широкий спектр принципиально различных алгоритмов. Их объединяет, как правило, только общий подход к обучению¹, но архитектура построения и принципы их работы различны. В задачах прогнозирования временных рядов с наилучшей стороны себя проявили рекуррентные и сверточные нейронные сети. Привлекательность рекуррентных нейронных сетей (RNN) состоит в том, что они потенциально умеют связывать предыдущую информацию с текущей. В отличие от полносвязных нейронных сетей, RNN используют внутреннюю память для обработки входных данных. Каждый блок RNN принимает текущие и предыдущие входные данные одновременно. Преимущество сверточных сетей заключается в количестве параметров, которое требуется обучить, по сравнению полносвязными или рекуррентными сетями при решении однотипной задачи. Идея применения CNN для прогнозирования временных рядов состоит в том, чтобы обучить фильтры, которые способны выделять определенные повторяющиеся шаблоны в данных, и использовать их для прогнозирования будущих значений. Благодаря многоуровневой структуре CNN могут хорошо работать с зашумленными данными, отбрасывая в каждом последующем слое шум и выделяя только значимые шаблоны, имея сходство, таким образом, с моделями, которые используют вейвлет-разложение².

¹ Алгоритм обратного распространения ошибки в основе которого находится градиентный спуск.

² Разложение временного ряда на его частотные компоненты.

2 Разработка модели прогнозирования индексов производства отечественной экономики с учетом межотраслевых связей

Экономические системы часто переключаются с одного режима на другой. Когда происходит такое переключение, распределение данных изменяется. В случае экономического цикла, макроэкономика периодически переключается с экспансии на рецессию и обратно, и динамика между этими двумя режимами различается. Однако эффект асимметрии свойственен не только макроэкономическим агрегатам самого высокого уровня. Учет этих явлений имеет основополагающее значение для прогнозирования. В данной части работы будет описана концепция собственной прогнозной модели на основе искусственных нейронных сетей, отвечающая особенностям отраслевой динамики и технологическим взаимосвязям между секторами.

2.1 Особенности отраслевой динамики и технологические взаимосвязи

В работе предпринимается попытка построить прогноз для трех индексов производства по структуре ОКВЭД: «Добыча топливно-энергетический полезных ископаемых»; «Производство кокса, нефтепродуктов и ядерных материалов»; «Химическое производство» (рисунок). Экономическая ситуация в этих секторах во многом определяет общую макроэкономическую динамику в России.

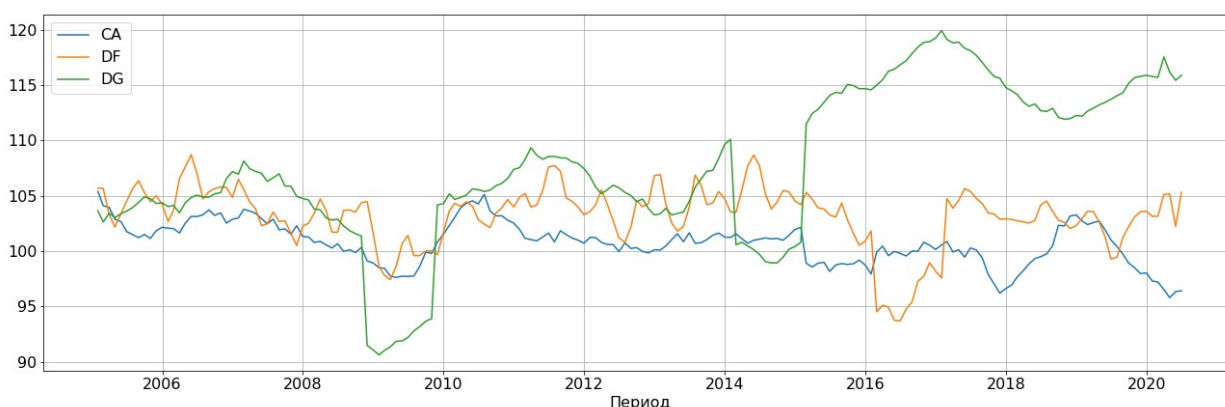


Рисунок 8 – индексы производства по структуре ОКВЭД, где CA - «Добыча топливно-энергетический полезных ископаемых»; DF – «Производство кокса, нефтепродуктов и ядерных материалов»; DG – «Металлургическое производство и производство готовых металлических изделий»

Предполагается, что для трех представленных выше отраслей возможны следующие технологические цепочки:

1. Добыча топливно-энергетических полезных ископаемых -> производство кокса, нефтепродуктов и ядерных материалов -> химическое производство;
2. Добыча топливно-энергетических полезных ископаемых -> химическое производство -> производство резиновых и пластмассовых изделий;

3. Добыча топливно-энергетических полезных ископаемых -> производство кокса, нефтепродуктов и ядерных материалов -> производство резиновых и пластмассовых изделий.

На стадии добычи топливно-энергетических полезных ископаемых производится сырье для отрасли, производящей кокс, нефтепродукты и ядерные материалы. В свою очередь химическая промышленность использует в качестве сырья в том числе, как продукты переработки второй отрасли, так и первой. На рисунке схематично изображены описанные межотраслевые связи с примерами технологических цепочек производства.

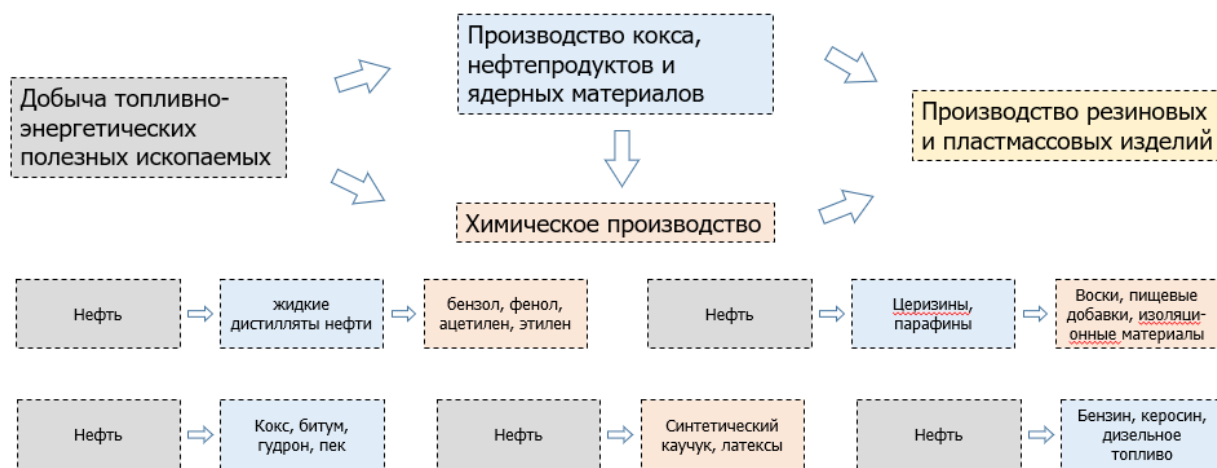


Рисунок 9 – направления технологических связей анализируемых отраслей с примерами технологических цепочек производства

Таким образом, выпуск отрасли добычи топливно-энергетических полезных ископаемых является сырьевым фактором отрасли производства кокса, нефтепродуктов и ядерных материалов, а также химической отрасли. Следовательно, структурная эконометрическая модель выпуска отрасли может включать выпуск другой отрасли, как экзогенный фактор. Однако для прогноза на период $t+1$ необходимо иметь $t+1$ значение выпуска вспомогательной отрасли. К тому же эконометрическая оценка подобной модели была бы невозможна ввиду проблемы эндогенности, возникающей вследствие существования обратной причинной связи.

Предлагаемый многомодельный подход позволяет разрешить данные проблемы: кроме непосредственно моделирования выпуска отрасли предлагаемая нейронная сеть одновременно моделирует сопутствующие факторы и строит прогноз вспомогательной отрасли, оказывающей воздействие на производственный процесс целевой отрасли, что и обеспечивает учет технологических связей. Это возможно за счет комбинирования различных отраслевых моделей в одну нейронную сеть.

2.2 Прогнозирование нелинейных временных рядов

Собственная прогнозная модель представляет собой комбинацию изначально независимых моделей, каждая из которых может быть использована в том числе, как самодостаточная для построения прогноза выпуска отрасли. Ввиду этого следует первоначально описать концепцию построения одной такой модели, а затем методологию их объединения.

В качестве основной архитектуры предполагается использовать сверточную нейронную сеть. Сверточные сети практически в современной форме появились в работах группы Яна ЛеКуна [29], и с тех пор успешно применяются для распознавания изображений, генерации звуковых последовательностей, предсказания временных рядов. Структура данных, подающихся на вход такой сети, несет в себе определенный смысл и влияет на конечный результат. Одним обучающим образцом, в случае анализа монохромного изображения, является его пиксельная сетка, имеющая двумерную размерность. В случае временных рядов аналогом пиксельного изображения, то есть одним обучающим образцом, выступает временное окно объясняющих факторов, имеющее также двумерную размерность: аналогом пикселей, вытянутых по горизонтали выступает вектор объясняющих переменных в один момент времени; аналогом пикселей, расположенных вдоль оси ординат, выступает вектор данных, представляющий изменение определенного фактора во времени. Идея применения сети состоит в том, чтоб изучать сверточными фильтрами периоды, которые представляют собой определенные повторяющиеся шаблоны и использовать извлеченную информацию о взаимозависимостях для прогнозирования будущих значений целевой переменной.

Концепции, описываемой далее модели взята из работы [10]. Отличия в теоретической части состоят в обобщении одномерного случая до многомерного и адаптации сети к решению задачи регрессии. В практической части мы предлагаем иную конфигурацию сверточных фильтров.

В формальном виде модель может быть представлена уравнением .

$$Y_t(\omega) = H(\omega, X_{N \times p}), \quad (13)$$

где

$H(\omega, X_{N \times p})$ – это нелинейная комбинация некоторого подмножества;

X – матрица исходных данных;

ω – вектор параметров;

p – количество факторов;

N – размер подмножества или иначе – плавающего временного окна.

Другими словами H – это функция, описывающая трансформацию $X_{N \times p}$ посредством серии сверточных фильтров. Имеет смысл сразу привести описание многослойной сети, поскольку в дальнейшем именно она будет использована в практической части работы. Тогда трансформацию исходных данных при помощи функции H можно описать следующим образом: если подмножество первоначальных данных X имеет размерность $N \times p$, где N – протяженность временного окна, определяющего один обучающий образец, p – это количество имеющихся факторов, то результат применения к матрице X набора фильтров первого слоя w_h^1 , где $h=1, \dots, M_1$ и $w_h^1 \in R^{p \times k \times 1}$, будет представлять собой матрицу³ A размерности $1 \times N - k + 1 \times M_1$, где k – это кол-во периодов, накрываемых фильтром, а элемент матрицы рассчитывается по формуле .

$$a^{l=1}(i, h) = (w_h^1 * x)(i) = \sum_{j=-\infty}^{\infty} w_h^1(j) x(i-j), \quad (14)$$

где

a_i – элемент результирующей матрицы свертки;

h – фильтр;

w_h^1 – вектор параметров;

N – протяженность временного окна, определяющего один обучающий образец;

p – это количество имеющихся факторов.

Кроме того, к матрице A применяется нелинейная функция $h(\cdot)$, т.н. функция активации, в результате чего $f^1 = h(a^1)$. Поскольку f^1 имеет размерность A , то на следующем слое каждый фильтр w_h^2 , где $h=1, \dots, M_2$, применяется к каждому столбцу матрицы A , в результате чего элемент матрицы A^2 , получаемой на следующем слое рассчитывается согласно формуле .

$$a^{l=2}(i, h) = (w_h^l * f^{l-1})(i) = \sum_{j=-\infty}^{+\infty} \sum_{m=1}^{M_1} w_h^l(j, m) f^{l-1}(i-j, m), \quad (15)$$

где

a_i – элемент результирующей матрицы свертки;

w_h^l – это фильтр h на слое l ;

³ Хотя по сложившейся терминологии искусственных нейронных сетей A стоит назвать тензором, поскольку здесь присутствует уже три пространства, мы будем называть A матрицей для того, чтоб избавить читателя от лишних многозначных терминов, поскольку размерность первого пространства всегда будет равна 1.

f^{l-1} – результат применения фильтров и функции активации на первом;

M_1 – это количество фильтров на первом слое.

К результату второго слоя также применяется активационная функция, в результате чего $f^2 = h(A^2)$. Затем процесс повторяется до последнего слоя L , на котором рассчитывается $f^L = h(A^L)$. К последней матрице применяется единичная свертка, сокращающая размерность f^L до $1 \times Z - k + 1$, где Z – длина f^{L-1} .

Полученный в результате вектор и является нелинейной комбинацией первоначальной матрицы X , обозначенный в формуле (10), как $H(\omega, X_{N \times p})$. Параметры функции H – это веса сверточных фильтров. Они определяются в результате оптимизации функция ошибки, которая представляет собой стандартную среднеквадратическую ошибку.

$$E(\omega) = \frac{1}{N} \sum_{i=0}^N \ddot{u}_i \ddot{u}_i \quad (16)$$

где

$E(\omega)$ – функция ошибки;

y_i – фактические данные;

\hat{y}_i – прогнозные данные;

ω – вектор параметров.

Оптимизация функции ошибки основана на градиентном спуске, в котором все параметры обновляются итерационно на основе градиента:

$$w_h^l(\tau+1) = w_h^l(\tau) - \eta \nabla E(w(\tau)), \quad (17)$$

где

w_h^l – вектор весов;

$\tau = 1, \dots, T$, T – количество итераций обучения;

η – скорость обучения.

Каждая итерация состоит из прохода вперед, в котором вычисляется прогнозируемый вектор \hat{y} и соответствующая ошибка $E(w)$, и прохода в обратном направлении, при котором рассчитывается вектор градиента $\nabla E(w(\tau))$, т.е. частные производные по каждому параметру, после чего происходит обновление всех параметров в соответствии с уравнением (14). Вектор градиентов вычисляется путем обратного распространения ошибки, что представляет собой итеративное применение правила

взятия производной от цепочно-вложенных функций. Вычисление частной производной по параметру $w_h^l(i, m)$ – i -му весу фильтра h на слое l происходит по формуле .

$$\frac{\partial E(w(\tau))}{\partial w_h^l(j, m)} = \sum_{\tau=1}^{N_i} \frac{\partial E(w(\tau))}{\partial f^l(i, h)} \frac{\partial f^l(i, h)}{\partial a^l(i, h)} \frac{\partial a^l(i, m)}{\partial w_h^l(i, m)} \quad (18)$$

где

$E(w(\tau))$ – функция ошибки;

w_h^l – вектор весов;

$\tau = 1, \dots, T$, T – количество итераций обучения;

h – фильтр;

l – слой;

a_i – элемент результирующей матрицы свертки;

Количество обучающих итераций T выбирается таким образом, чтобы обеспечить сходимость по ошибке. Следуя [10], для собственных расчетов используется оптимизатор Адам. Этот метод вычисляет адаптивные скорости обучения для каждого параметра, сохраняя экспоненциально убывающее среднее значение прошлых градиентов и квадратов градиентов, и использует их для обновления параметров. Адаптивная скорость обучения позволяет градиентному спуску находить минимум более точно.

Схематично процесс построения прогноза для отдельной отрасли при помощи описанной модели представлен на рисунке .

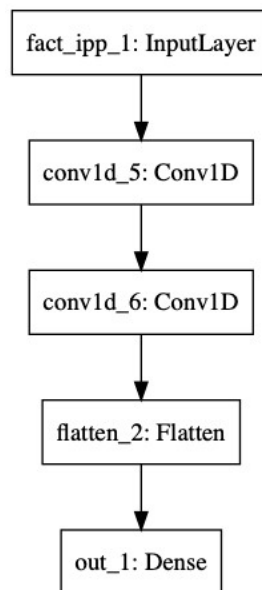


Рисунок 10 – схема построения прогноза для отдельной отрасли при помощи сверточной глубокой сети

Входные данные – это временное окно прогнозируемого временного ряда (индекса производства), в случае безусловного прогноза и временное окно матрицы экзогенных

факторов в случае условного прогноза. Первый и второй слои – это комбинации сверточных фильтров, производящих аффинные преобразования с входными данными. Выходные данные – это прогноза на $t+1$ период времени.

Концепция многомодельного подхода предполагает объединение индивидуальных моделей, предназначенных для прогнозирования отдельных индексов производства в единую систему по следующей схеме (рисунок). Идея объединения трех моделей состоит в том, что отрасли связаны производственными связями и прогноз одной отрасли влияет на прогноз другой. В рамках одномодельного подхода так же можно учесть влияние смежной отрасли путем включения в перечень факторов предложения выпуска отрасли, производящей ресурсы производства. Одномодельный подход предполагает использовать в качестве входных данных только факторы, непосредственно влияющие на объемы производства (как со стороны спроса, так и со стороны предложения) определенной отрасли. В свою очередь многомодельный подход позволяет учитывать не только выпуск смежной отрасли и его лаги, но и факторы производства вспомогательной отрасли в предположении о том, что изменение факторов спроса/предложения такой отрасли через некоторый временной лаг приведет к изменению объемов производства целевой отрасли и нейронная сеть сможет использовать подобные шаблоны в данных для построения прогноза. Наличие модели, имеющие подобные возможности представляется необходимым в текущих условиях, когда нарушение глобальных цепочек производства и распределения ресурсов приводит к возникновению кризисных ситуация в отраслях, связанных как непосредственно между собой, так и опосредованно. Это возможно благодаря объединению индивидуальных прогнозных моделей после преобразования исходных данных каждой отрасли через свою серию сверточных фильтров, которые, как предполагается определяют повторяющиеся шаблоны в данных, позволяющие построить прогноз на следующие период времени.

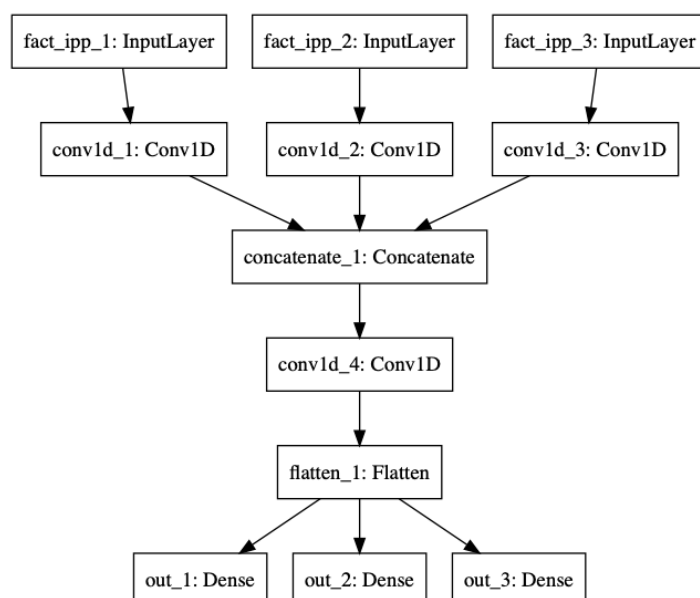


Рисунок 11 – концепция многомодельного подхода на основе индивидуальных прогнозных моделей

Слияние информационных потоков, поступающих от трех моделей, происходит в блоке «Concatenate_1», после чего получившаяся матрица проходит через блок снижения размерности «Conv1d_4». Полученные в результате данные поступают на вход в каждый из трех блоков последнего уровня, которые отвечают за независимую минимизацию отклонения ошибки прогноза для каждой отрасли, входящей в производственную цепочку. При этом в модели предусмотрено, как принудительное разделение информационных потоков от трех моделей по важности, так эндогенное. В принудительном случае ранжирование информационных потоков осуществляется на этапе конкатенации перед блоком «Concatenate_1», что происходит в соответствии с экзогенно заданными весовыми коэффициентами. В случае эндогенного разделения потоков информации, взвешиваются только конечные функции ошибки, где максимальный вес придается ошибке, которая рассчитывает отклонение прогноза от факта для целевой отрасли. Поскольку в конечном итоге минимизируется функция ошибки, фактически состоящая из трех отдельных функций ошибки, предполагается, что в соответствии со значимостью отдельной ее составляющей, в процессе обучения будет происходить настройка параметров модели наилучшим образом. Выбор устройства конечной модели зависит от эмпирических экспериментов, которые будут описаны следующей части работы. При этом важно обращать внимание не только на метрики качества прогноза на тестовой выборке, но и на распределение весовых коэффициентов, поскольку короткой истории наблюдений сеть может переобучаться на тренировочной части, «запомнив» экономически не интерпретируемые шаблоны в данных, случайным образом приводящие к высокому качеству прогноза на тестовой выборке.

2.3 Определение объясняющих факторов и содержательные гипотезы

Для определения перечня объясняющих факторов и формулирования гипотез исследования для каждой из исследуемых отраслей проводится построение систем уравнений спрос-предложение.

Уравнения (19) - (21) представляют собой систему одновременных уравнения спрос-предложение, описывающих равновесное состояние в отрасли добычи топливно-энергетических полезных ископаемых.

$$Y_t^{d_1} = \beta_0 + \beta_1 P_t^{oil} + \beta_2 P_t^{gaz} + \beta_3 P_t^{coal} + \beta_4 H_t + \beta_5 Inc_t^M + \beta_6 Inc_t^H + \beta_7 exp_t^c + \beta_8 Er_t + \beta_9 \mathfrak{T}_t + \eta \quad (19)$$

$$Y_t^{s_1} = \theta_0 + \theta_1 P_t^{oil} + \theta_2 P_t^{gaz} + \theta_3 P_t^{coal} + \theta_4 I_t + \theta_5 W_t + \theta_6 Tr_t + \theta_7 exp_t^p + \theta_8 OPEC_t + \theta_9 G_t + \vartheta \quad (20)$$

$$Y_t^{d_1} = Y_t^{s_1} + \psi \quad (21)$$

где

$Y_t^{d_1/s_1}$ – предложение/спрос отрасли;

P_t^{oil} – средние биржевые цены крупнейших производителей нефти;

P_t^{gaz} – средние биржевые цены крупнейших производителей газа;

P_t^{coal} – средние биржевые цены крупнейших производителей угля;

H_t – производство теплоэнергии;

Inc_t^M – Доход производителей кокса, нефтепродуктов, ядерных материалов

Inc_t^H – Доход химической отрасли

Exp_t^c – ожидания потребителей

\mathfrak{T}_t – импорт энергоресурсов;

Er_t – обменный курс рубля;

I_t – инвестиции в отрасль;

W_t – заработная плата в отрасли;

Tr_t – издержки транспортировки;

Exp_t^p – ожидания производителей

$OPEC_t$ – ограничения ОПЕК;

G_t – государственные субсидии;

η, ϑ, ψ – нормально распределенные ошибки с нулевым математическим ожиданием и постоянной дисперсией.

Предложение/спрос отрасли являются эндогенными переменными. Средние биржевые цены крупнейших производителей нефти – эндогенная переменная, которая

отражает стоимость продукции, производимой отраслью. В качестве прокси для данной переменной может выступать средняя спотовая цена нефти марок Brent, Dubai и West Texas Intermediate и Ural. Средние биржевые цены крупнейших производителей газа – эндогенная переменная, которая отражает стоимость продукции, производимой отраслью. Средние биржевые цены крупнейших производителей угля – эндогенная переменная, стоимость продукции, производимой отраслью. Гипотеза: все цены оказывает положительное влияние на предложение, отрицательное на спрос. Указанные цены могут быть заменены одной в рамках моделирования, прокси для которой является индекс товарного топлива, который включает в себя индексы цен на сырую нефть, природный газ и уголь. Производство теплоэнергии – фактор спроса. Угольные, Газовые ТЭЦ являются крупными потребителями производимой отраслью продукции. Гипотеза: снижение объемов теплоэнергии, производимой данными видами теплоэлектростанций приводит к сокращению спроса на продукцию отрасли. Доход производителей кокса, нефтепродуктов, ядерных материалов – фактор спроса. Гипотеза: увеличение дохода основных потребителей приводит к росту спроса на производимую отраслью товар. В качестве прокси для данного показателя может использоваться индекс производства отрасли. Доход химической отрасли – фактор спроса. Гипотеза: увеличение дохода отрасли приводит к росту спроса на производимый отраслью товар. Ожидания потребителей – фактор спроса. Гипотеза: оптимистичные ожидания потребителей приводят к увеличению спроса на продукцию отрасли. Данный показатель может быть непосредственно измерен по данным Росстата как индекс предпринимательской уверенности организаций, обеспечивающих электрической энергией, газом и паром; кондиционированием воздуха. Импорт энергоресурсов – фактор спроса. Гипотеза: увеличение объемов импорта при прочих равных означает увеличение спроса на продукцию отрасли. Обменный курс рубля – фактор в данном случае спроса/предложения. Гипотеза: в результате ослабления отечественной валюты относительная стоимость продукции отрасли на международном рынке снижается, что вызывает рост спроса со стороны внешних рынков сбыта. Инвестиции в отрасль – фактор предложения. Гипотеза: увеличение инвестиций в отрасль влечет расширение производств и, следовательно, увеличение предложения. Заработная плата в отрасли – фактор предложения. Увеличение заработной платы – сигнал для рынка труда в необходимости расширения штата специалистов в данной отрасли, приток рабочей силы обеспечивает задействование большего количества ресурсов производства, в результате чего возможно повышение предложения отрасли. Подобная картина особенно хорошо видна на примере нефтяного сектора в России, где уровень заработных плат является самым высоким среди прочих отраслей

промышленности. Несмотря на то, что заработная плата – издержки производства, в данном случае представляется, что уровень заработной платы в отрасли может иметь отрицательный знак в уравнении предложения скорее в странах с сильными профсоюзами, способными повысить уровень заработной платы для работников вне зависимости от рентабельности производства. Чрезмерное повышение может привести к тому, что часть фирм становятся нерентабельными и сокращают производство, что снижает предложение. Издержки транспортировки – фактор предложения. Гипотеза: при увеличении издержек транспортировки некоторая часть производителей становятся нерентабельными, предложение сокращается. Ожидания производителей – фактор предложения. Оптимистичные ожидания производителей приводят к увеличению предложения на продукцию отрасли. В данном случае показатель может быть напрямую измерен при помощи индекса предпринимательской уверенности по добыче полезных ископаемых по данным Росстата. Ограничения ОПЕК – фактор предложения. Гипотеза: ожидания сокращения спроса странами ОПЕК вынуждают стран участниц сокращать добычу нефти, что приводит к увеличению цены на нефть и увеличению предложения отечественной нефти (предложение отрасли увеличивается).

Добыча сырой нефти Организацией стран-экспортеров нефти (ОПЕК) является важным фактором, влияющим на цены на нефть. Эта организация стремится активно управлять добычей нефти в странах-членах, устанавливая производственные цели. Исторически сложилось так, что цены на сырую нефть росли во время снижения планов ОПЕК по добыче. Стоит заметить, что страны-члены ОПЕК производят около 40 процентов мировой нефти. Кроме того, экспорт нефти ОПЕК составляет около 60 процентов от общего объема торговли нефтью на международном уровне. Из-за такой доли рынка действия ОПЕК могут существенно влиять на мировые цены на нефть.

Последний фактор - государственные субсидии. Гипотеза: Субсидии в отрасль приводят к расширению производств и увеличению предложения.

Уравнения (22) - (24) представляют собой систему одновременных уравнений спрос-предложение, описывающих равновесное состояние в отрасли производства кокса, нефтепродуктов и ядерных материалов.

$$Y_t^{d_2} = \beta_0 + \beta_1 P_t^{Coke} + \beta_2 P_t^{Petr} + \beta_3 P_t^{NM} + \beta_4 H_t + \beta_5 Inc_t^H + \beta_6 Inc_t^M + \beta_7 Inc_t^P + \beta_8 Er_t + \beta_9 \Im_t + \eta \quad (22)$$

$$Y_t^{s_2} = \theta_0 + \theta_1 P_t^{Coke} + \theta_2 P_t^{Petr} + \theta_3 P_t^{NM} + \theta_4 I_t + \theta_5 W_t + \theta_6 Tr_t + \theta_7 exp_t^p + \theta_8 G_t + \vartheta \quad (23)$$

$$Y_t^{d_2} = Y_t^{s_2} + \psi \quad (24)$$

где

Y_t^{d/s_2} – предложение/спрос отрасли;

P_t^c – средние цены крупнейших производителей кокса;

P_t^{gaz} – средние цены крупнейших производителей нефтепродуктов;

P_t^{coal} – средние цены крупнейших производителей ЯМ;

H_t – производство электроэнергии/теплоэнергии;

Inc_t^H – доходы химической отрасли;

Inc_t^M – доходы металлургической отрасли;

Inc_t^P – доходы населения;

Er_t – обменный курс рубля;

$I m_t$ – импорт нефтепродуктов;

I_t – инвестиции в отрасль;

W_t – заработная плата в отрасли;

$E x p_t$ – ожидания производителей;

$T r_t$ – издержки транспортировки;

G_t – государственные субсидии;

β, θ – векторы оцениваемых коэффициентов;

η, ϑ, ψ – нормально распределенные ошибки с нулевым математическим ожиданием и постоянной дисперсией.

Предложение и спрос отрасли – это эндогенные переменные. Средние цены крупнейших производителей кокса; средние цены крупнейших производителей нефтепродуктов; средние цены крупнейших производителей ядерных материалов (ЯМ) – эндогенные переменные, которые отражают стоимость продукции, производимой отраслью. Гипотеза: положительное влияние на предложение, отрицательное на спрос. Производство электроэнергии/теплоэнергии – экзогенная переменная, являющаяся фактором спроса. Мазутные ТЭЦ и АЭС являются крупными потребителем производимой отраслью продукции. Снижение объемов теплоэнергии, производимой данными видами теплоэлектростанций приводит к сокращению спроса на продукцию отрасли.

Доходы химической отрасли; доходы металлургической отрасли – экзогенные переменные, являющиеся факторами спроса. Гипотеза: увеличение дохода основных потребителей приводит к росту спроса на производимы отраслью товар. Доходы населения – также фактор спроса. Гипотеза: доход потребителей продукции (прежде всего население предъявляет спрос на горюче-смазочные материалы, которые являются потребительскими продуктами, производимыми отраслью) снижается, падает их покупательная способность, спрос сокращается. Обменный курс рубля – фактор,

оказывающий одновременно влияние, как на спрос, так и на предложение. Поскольку в первом случае способен оказывать влияние на покупательскую способность населения, во втором случае на конкурентное преимущество производителя на мировом рынке. Гипотеза: в результате ослабления отечественной валюты относительная стоимость продукции отрасли на международном рынке снижается, что вызывает рост спроса со стороны внешних рынков сбыта. Кроме того, показателем покупательских настроений населения и одновременно стабильности функционирования банковской системы может являться изменение средств на счетах физических и юридических лиц. Импорт нефтепродуктов – фактор спроса. Увеличение объемов импорта при прочих равных означает увеличение спроса на продукцию отрасли, но при этом может происходить замещение использования отечественного продукта импортным.

Инвестиции в отрасль – фактор предложения. Гипотеза: увеличение инвестиций в отрасль влечет расширение производств и, следовательно, увеличение предложения. В качестве прокси данного показателя могут выступать объемы кредитов и прочие размещенные средства, предоставленные нефинансовым организациям. Заработная плата в отрасли – фактор предложения. С одной стороны, если рассматривать данный фактор на уровне производственной функции отдельной организации, то из классической теории фирмы следует, что заработная плата работников – это издержки отрасли, и, соответственно, при прочих равных увеличение заработной платы должно вызывать снижение объемов производства. Однако на уровне отрасли, в реалиях отечественной экономики и, в частности, для рассматриваемых отраслей, представляется действенным другой механизм: увеличение средней заработной платы в отрасли – это конкурентная борьба отрасли с другими за квалифицированную рабочую силу. Необходимость, соответственно, в задействовании большего количества ресурсов рабочей силы возникает, когда отрасль стремится расширить производство. Гипотеза, которую можно на основании этого выдвинуть: увеличение заработной платы в отрасли приводит к притоку рабочей силы и специалистов и за счет мобилизации ресурсов растет предложение отрасли.

Ожидания производителей – фактор предложения. Оптимистичные ожидания производителей приводят к увеличению предложения на продукцию отрасли. В данном случае ожидания производителей могут быть достаточно точно измерены при помощи индекса предпринимательской уверенности организаций, обеспечивающих электрической энергией, газом и паром; кондиционированием воздуха, публикуемым Росстатом.

Издержки транспортировки – фактор предложения. Гипотеза: при увеличении издержек транспортировки некоторая часть производителей становятся нерентабельными,

предложение сокращается. В качестве прокси для данного фактора могут выступать тарифы РЖД по прейскуранту 20-01 на перевозку товаров, производимых отраслью. В первую очередь, к таким товарам относятся нефтепродукты, перевозимые в больших объемах по железной дороге. Государственные субсидии – фактор предложения. Субсидии в отрасль приводят к расширению производств и увеличению предложения.

Уравнения (25) - (27) представляют собой систему одновременных уравнения спрос-предложение, описывающих равновесное состояние в отрасли производства кокса, нефтепродуктов и ядерных материалов.

$$Y_t^{d_3} = \beta_0 + \beta_1 P_t^{Fer} + \beta_2 P_t^{FP} + \beta_6 Inc_t^M + \beta_7 Inc_t^P + \beta_8 Er_t + \beta_9 \Im_t + \beta_9 Inf\ l_t + \eta \quad (25)$$

$$Y_t^{s_3} = \theta_0 + \theta_1 P_t^{Fer} + \theta_2 P_t^{FP} + \theta_3 P_t^{Oil} + \theta_2 I_t + \theta_3 W_t + \theta_5 Tr_t + \theta_7 E\ xP_t^P + \theta_7 G_t + \vartheta \quad (26)$$

$$Y_t^{d_3} = Y_t^{s_3} + \psi \quad (27)$$

где

$Y_t^{d_3/s_3}$ – предложение/спрос отрасли,

P_t^{Fer} – средние цены неорганических удобрений;

P_t^{FP} – индекс цен фармацевтической промышленности;

P_t^{Oil} – стоимость нефти;

Inc_t^M – доходы металлургической отрасли;

Inc_t^P – доходы населения;

Er_t – обменный курс рубля;

Im_t – импорт химической продукции;

$Inf\ l_t$ – уровень потребительской инфляции;

I_t – инвестиции в отрасль;

W_t – заработная плата в отрасли;

Tr_t – издержки транспортировки;

$E\ xP_t^P$ – ожидания производителей

G_t – государственные субсидии;

β, θ – векторы оцениваемых коэффициентов;

η, ϑ, ψ – нормально распределенные ошибки с нулевым математическим ожиданием и постоянной дисперсией.

Предложение и спрос химической отрасли являются эндогенными переменными, равновесные значения которых определяются внутри системы в каждый момент времени.

Забегая вперед, стоит заметить, что наблюдать данный фактор в действительности невозможно, поскольку статистика стоимостных объемов производства химической отрасли не доступна. В качестве прокси показателя может выступать индекс химического производства, публикуемый Росстатом [84]. Средние цены неорганических удобрений – эндогенная переменная, которая отражает стоимость продукции, производимой отраслью. Гипотеза: положительное влияние на предложение, отрицательное на спрос. Учитывая, что экспорт удобрений, произведенных отечественной экономикой, на мировом рынке занимает подавляющую долю, то можно предположить, что данный ценовой фактор в существенной степени оказывает влияние на равновесное положение в отрасли. Для эмпирических расчетов в качестве прокси для данного фактора могут выступать биржевые цены различных видов удобрений: фосфатные удобрения, долларов США за метрическую тонну, Indexmundi [88]; калийные удобрения, долларов США за метрическую тонну, Indexmundi [88]. Индекс цен фармацевтической промышленности – эндогенная переменная, определяющая равновесное положение отрасли. В свою очередь, фармацевтическая промышленность является крупной составляющей химической промышленности. Гипотезу для обозначенных стоимостных факторов можно сформулировать следующим образом: стоимость продукции, производимой отраслью; положительное влияние на предложение, отрицательное на спрос.

Стоимость нефти – фактор предложения, отражающий часть издержек отрасли. Частично переработанная нефть и сырая нефть – являются фактором производства ряда продуктов, производимых отраслью. В первую очередь это относится к производству пластмасс и полимеров. Гипотеза: с ростом цены, предложение отрасли сокращается, спрос увеличивается. Доходы металлургической отрасли – фактор спроса. Металлургическая отрасль является потребителем ряда продуктов, производимых отраслью. Соответственно, увеличение доходов металлургов приводит к расширению производства и повышению спроса на продукцию химической отрасли.

Доходы населения – фактор предложения. Население предъявляет спрос на широкий перечень продукции, производимой отраслью: бытовая химия, фармацевтическая продукция, лакокрасочные материалы и прочее. Гипотеза: снижение доходов населения приводит к снижению спроса на продукцию широкого потребления, производимую отраслью. Обменный курс рубля – фактор, оказывающий влияние одновременно на сторону формирования спроса и предложения. В случае спроса: относительное удешевление отечественной валюты к резервным мировым валютам приводит к росту общего уровня цен и снижению покупательной способности населения. Следовательно, спрос на продукцию отрасли может снизиться. В случае предложения:

относительное удешевление отечественной валюты может привести к тому, что относительная стоимость продукции отрасли на международном рынке снизится, что вызовет рост внешнего спроса.

Уровень потребительской инфляции – фактор спроса. Данный фактор непосредственно влияет на покупательную способность населения. Рост инфляции может быть вызван, как волатильностью отечественной валюты, так и внутренними процессами. При этом инфляция непродовольственных товаров может быть фактором, оказывающим влияние на предложение отрасли. Рост стоимости факторов производства может привести к тому, что часть предприятий окажется нерентабельными и предложение отрасли сократится. Фактором спроса также является импорт химической продукции. Гипотеза: увеличение объемов импорта при прочих равных означает увеличение спроса на продукцию отрасли.

Инвестиции в отрасль – фактор предложения. Гипотеза: увеличение инвестиций в отрасль влечет расширение производств и, следовательно, увеличение предложения. Заработная плата в отрасли, как и в предыдущих двух случаях, нами будет рассматриваться, как фактор предложения, положительно влияющий на выпуск. Предположительно, увеличение заработной – сигнал для рынка труда в необходимости расширения штата специалистов в данной отрасли для расширения производства. Издержки транспортировки – фактор предложения. Гипотеза: при увеличении издержек транспортировки некоторая часть производителей становятся нерентабельными, предложение сокращается. И последний фактор предложения – государственные субсидии. Гипотеза: государственная поддержка предприятий отрасли приводит к увеличению предложения.

Заметим, что параметры уравнений трех систем сверхидентифицируемы. Известно, что параметры системы одновременных уравнений однозначно идентифицируемы, если количество экзогенных переменных в системе минус количество эндогенных переменных в уравнении равно количеству эндогенных переменных в системе минус количество эндогенных переменных в уравнении. Подобное правило должно выполняться для каждого уравнения. Другими словами, параметры системы являются идентифицируемыми, если существует возможность восстановления коэффициентов структурной формы уравнений по коэффициентам приведенной формы. Коэффициенты отдельного структурного уравнения не могут быть восстановлены на основании коэффициентов приведенной формы, если в это уравнение входят все эндогенные переменные [30]. Если в левой части приведенного уравнения находится одна единственная эндогенная переменная, то условие идентифицируемости можно

сформулировать следующим образом: «количество экзогенных переменных в системе, не включенных в определенное структурное уравнение, должно быть не меньше количества эндогенных переменных в правой части этого уравнения» [30].

В случае имеющихся систем одновременных уравнений выполняется следующее: разность числа экзогенных переменных в каждой системе с каждым уравнением больше, чем в соответствующих системах разность числа эндогенных переменных в системе и каждом уравнении. Следовательно, параметры всех систем являются сверх идентифицируемыми и нельзя использовать приведенные уравнения для оценки параметров. В таком случае можно использовать двухшаговый метод наименьших квадратов. Однако двухшаговый метод наименьших квадратов позволит проверить статистическую значимость наличия только линейной связи между независимыми переменными. В то время как в нейросетевой модели, которая предполагается в качестве основной прогностической модели в данном исследовании, происходит нелинейное преобразование всех независимых переменных. Нейронная сеть в задаче регрессии предполагает линейную связь с объясняемой переменной только на последнем уровне нейронной сети. Однако признаки последнего уровня – это нелинейная комбинация всех признаков. При этом исключение одного признака из нейронной сети, который не имеет статистически значимой линейной связи, что подтверждается незначимым коэффициентом перед фактором, полученным в результате оценки двухшаговым методом наименьших квадратов, может привести к снижению качества прогноза. Подобная ситуация как раз может быть вызвана тем, что линейной статистической связи, действительно не наблюдается между фактором и объясняемой переменной, однако нелинейная связь может существовать.

Можно заметить, что в правых частях уравнений систем наблюдаются пересекающиеся факторы, а также экзогенные факторы, которые в других уравнениях являются эндогенными. На основании этого можно сделать вывод, что рассматриваемые отрасли имеют тесную взаимосвязь. Следовательно, состояние в одной отрасли влияет на экономическое положение в другой отрасли.

Гамильтон [31] утверждает, что, хотя факторные модели являются точным описанием состояния экономики, их полезность в качестве инструментов прогнозирования ограничена, если они включают в себя вводящие в заблуждение индикаторы, или если поведение встроенных индикаторов со временем изменилось. По этой причине мы предполагаем, что точность прогноза на основе оцененных эластичностей приведенных систем одновременных уравнений может быть ниже, чем точность прогноза, построенного на основе аналогичных данных с использованием

нейронной сети. Исходя из теоретического анализа нейронных сетей, поведенного в первой главе исследования, можно сделать следующие вывод относительно сравнительных преимуществ сетей с системами одновременных уравнений:

1) Нейронная сеть не ограничена функциональными взаимоотношениями системы одновременных уравнений и, соответственно, имеет большую гибкость в использовании прогностических качеств факторов.

2) Нейронная сеть позволяет учитывать изменение влияния фактора на целевую переменную во времени, поскольку топология глубокой нейронной сети подразумевает учет пороговых эффектов. При этом вовсе не обязательно проводить переобучение нейронной сети. В том случае, если в числе обучающих образцов, будут содержаться паттерны с противоположным влиянием одного фактора на целевую переменную, внутри сети будет создано две ветви нейронов, каждая из которых активируется только в том случае, если встречается подобный паттерн поведения на тестовой выборке.

3) Система одновременных уравнений предполагает равновесие на рынке. Соответственно прогноз, построенный при помощи приведенного уравнения системы, будет предполагать нахождение в равновесии в каждый момент времени. В таком случае при получении прогноза объема автоматически происходит прогнозирование цены (при подстановке объема в одно из уравнений системы). Аналогично, верно, и для цены. Однако, если принимать за истину кейнсианскую и неокейнсианскую теории общего экономического равновесия, то возникает необходимость прогнозировать состояние экономики или в данном случае отрасли, которая находится только в частичном равновесии или вовсе отклоняется от него. Таким образом, прогноз при помощи параметров из системы одновременных уравнений может быть оправдан, если предполагается равновесие в экономике в каждый момент времени. Нейронная сеть не прогнозирует равновесное состояние. Цена является одним из эндогенных факторов, влияние которого на эндогенную переменную может меняться со временем.

При разработке собственной прогнозной модели мы исходили из того, что отрасли связаны между собой. К примеру, на стадии добычи топливно-энергетических полезных ископаемых производится сырье для отрасли, производящей кокс, нефтепродукты и ядерные материалы. В свою очередь химическая промышленность использует в качестве сырья в том числе, как продукты переработки второй отрасли, так и первой. Таким образом, выпуск отрасли добычи топливно-энергетических полезных ископаемых является сырьевым фактором отрасли производства кокса, нефтепродуктов и ядерных материалов, а также химической отрасли. Следовательно, структурная эконометрическая модель выпуска отрасли может включать выпуск другой отрасли, как экзогенный фактор. Однако

оценить такую систему уравнений эконометрически и получить из нее прогноз крайне сложно, поскольку в каждом уравнении присутствует проблема эндогенности, к тому же остается нерешенной проблема нелинейности. Предлагаемая система нейронных сетей позволяет разрешить данные проблемы: прогноз строится одновременно для трех отраслей таким образом, что прогноз одной отрасли влияет на формирование прогноза в другой. Прогнозная модель в итоге представляет собой комбинацию индивидуальных моделей, каждая из которых может быть использована как самостоятельная. Индивидуальная модель в системе – это сверточная нейронная сеть. Ее целевая переменная – индекс производства отрасли, а объясняющие переменные – это факторы спроса и предложения данной отрасли. Для определения перечня объясняющих факторов и формулирования соответствующих им гипотез для каждой из исследуемых отраслей были построены теоретические системы уравнений «спрос-предложение».

3 База данных эмпирического исследования

3.1 Описание структуры базы данных

База данных исследования содержит месячные и квартальные данные с января 2005 года по август 2020 гг. В качестве зависимых переменных в моделях выступают индексы промышленного производства: индекс производства добычи топливно-энергетический полезных ископаемых; индекс производства добычи топливно-энергетический полезных ископаемых; индекс производства химической отрасли. Для того, чтоб иметь максимально возможную длину временных рядов применялся пересчет индексов промышленного производства из ОКВЭД2 в ОКВЭД1 по методологии ИЭП [32]. В качестве независимых переменных, в соответствии с описанными в предыдущем разделе гипотезами, выступают факторы из отраслевого, реального и финансового блоков. К отраслевым факторам относятся индивидуальные показатели со стороны спроса и предложения определенной отрасли. К факторам реального блока относятся показатели сырьевого рынка, оказывающие прямое или опосредованное влияние на все анализируемые отрасли. К факторам финансового блока относятся общие для всех отраслей факторы финансового и банковского сектора, как правило, характеризующие сторону спроса. Все данные приводились к месячным значениям. Высокочастотные данные, такие как валютный курс и биржевые цены сырьевых товаров, были усреднены. Квартальные данные перед использованием в эмпирической части работы были интерполированы до месячных данных в случае необходимости приведения экзогенных факторов к единой размерности. В качестве источников преимущественно использовались данные Росстата [33], Банка России [34]; международная статистика была взята с ресурсов IMF [35], EIA [36].

3.2 Предобработка данных для нейросетевого моделирования

Предобработка данных для нейросетевого моделирования включает в себя следующие этапы: проверка на стационарность; приведение рядов к стационарному виду в случае наличия единичного корня; масштабирование; применение метода главных компонент для факторов с высокой корреляцией; приведение данных к необходимой размерности для обучения нейронной сети.

Проверка на стационарность и приведение к стационарности рядов необходимо для возможности выполнения пункта три: масштабирования переменных. Масштабирование, в данном случае, это приведение всех данных в диапазон от -1 до 1. Из каждого наблюдения вычитается математическое ожидание и разность делится на среднеквадратическую ошибку. Стандартизация переменных необходима для получения более гладкой поверхности функции ошибки, что обеспечивает нормальную работу градиентного спуска в процессе подбора параметров нейронной сети. Следовательно, для

того, чтоб существовала возможность стандартизировать все переменные необходимо по каждой переменной иметь математическое ожидание дисперсию постоянную на подвыборках. При этом необходимо проводить стандартизацию не только объясняющих переменных, но и объясняемой, поскольку значительно отличающийся масштаб объясняемой переменной может также приводить к поверхности функции ошибки с резкими обрывами и перепадами, что затруднит нахождение локального минимума функции, удовлетворяющего критериям прогнозной ошибки. Поскольку для чистоты эксперимента математическое ожидание и дисперсия для приведения объясняемой переменной к ее естественному масштабу на тестовой выборке, при построении прогнозов с применением уже обученной сети должны применяться из обучающей выборки, объясняемая переменная также должна быть стационарна.

В таблице приведены результаты проверки экзогенных и эндогенных факторов для модели каждой отрасли на наличие единичного корня.

Таблица 1 – Проверка на стационарность при помощи теста Дики-Фуллера

| Название переменной | Статистика теста | р-значение | Критическое значение (10%) | Результат проверки |
|---------------------|------------------|------------|----------------------------|---------------------------------------|
| CA_ch | - 4,15 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| inconf_r | - 5,45 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| inconf_e | - 9,28 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ps_ca | - 6,07 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| DF_ch | - 4,56 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ps_gasoline | - 7,09 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ps_m | - 5,93 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ps_ew | - 6,32 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| DG_ch | - 3,64 | 0,01 | - 2,58 | единичных корней нет, ряд стационарен |
| ps_oil | - 6,32 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| dap | - 4,80 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| potass | - 5,00 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| phosphate | - 4,35 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| nomex | - 3,86 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |

Продолжение таблицы

| Название переменной | Статистика теста | p-значение | Критическое значение (10%) | Результат проверки |
|---------------------|------------------|------------|----------------------------|---------------------------------------|
| Brent | - 6,20 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ipp_nonprod | - 9,07 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| ipp_prod | - 4,70 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| cost_e | - 6,40 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| credit | - 5,95 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| deposit_firm | - 4,16 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| deposit_peop | - 6,32 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| m2_seas_adj | - 4,99 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| Spare_Capacity | - 4,53 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| WTI_Real_Price | - 4,95 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| Saudi_Prod_Change | - 4,15 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |
| WTI_Prod_Change | - 4,66 | 0,00 | - 2,58 | единичных корней нет, ряд стационарен |

Из результатов проверки на наличие единичного корня следует, что все рассматриваемые ряды являются стационарными. При этом, часть показателей была предварительно очищена от линейного тренда и сезонности при помощи метода X12-ARIMA. Следовательно, полученные в результаты временные ряды могут быть нормированы, а объясняемая переменная для проверки прогностических качеств модели, может быть восстановлена в естественную шкалу при помощи оценок математического ожидания и дисперсии, полученных на тренировочной выборке.

Приведение данных к определенной размерности обусловлено применением специальных архитектур нейронных сетей, который предназначены для обработки временных рядов. Для всех используемых архитектур: двух разновидностей рекуррентных нейронных сетей и сверточной нейронной сети необходима одна форма подачи обучающих образцов и, соответственно, такая же форма тестовых образцов для проверки качества работы нейронной сети. Форма входных данных в нейронной сети в том числе является гиперпараметром, поскольку определяет явно какое количество лагов объясняемой и объясняющих переменных будет участвовать в построении прогноза на один шаг. В результате тестирования однофакторных моделей в разделе 2.2

Прогнозирование нелинейных временных рядов было определено, что оптимальное окно входного паттерна или образца данных составляет от трех месяцев до шести. При этом важно отметить, что рекуррентная нейронная сеть имеет к тому же и долгосрочную память, поскольку умеет связывать паттерны между собой – каждый нейрон рекуррентной сети, кроме обучающего образца принимает на вход вектор, являющийся результатом обработки предыдущего обучающего образца тем же нейроном. Для того, чтоб позволить сверточной сети также обнаруживать и учитывать закономерности во взаимодействии факторов в отдаленной истории наблюдений, было принято решение для сверточной сети использовать окно в 2 квартала или шесть месяцев, а для рекуррентных архитектур окно в 1 квартал или три месяца. Подобное решение так же обусловлено необходимостью экономии общего числа параметров для обучения, поскольку число наблюдений во времени относительно не велико, а большое низкое соотношение числа наблюдений к числу обучающихся параметров может привести к переобучению и неустойчивым прогнозам. Более широкое временное окно предназначено для сверточной сети, а не рекуррентной, поскольку, если сопоставить одно сверточное ядро с $2 \times n$ размерным фильтром при прочих равных с рекуррентным нейроном, то сверточное ядро будет содержать меньшее количество обучаемых параметров. При прочих равных означает, что размерность входного паттерна должны быть одинаковой. $2 \times n$ размерный фильтр – это минимальная размерность стандартного фильтра для обработки временных рядов, где два – это сколько строк в матрице пространства признаков фильтр сворачивает в один момент времени, а n – это количество экзогенных переменных. Вторая размерность является всегда фиксированной и равной числу признаков, поскольку предполагается, что каждый фактор вносит свой некоторый индивидуальный вклад в аппроксимацию целевой переменной. В случае использования усеченного фильтра одни и те же веса будут использоваться для разных факторов при составлении линейной комбинации признаков перед функцией активации. Таким образом, временное окна входящего паттерна для сверточной нейронной сети можно считать максимально допустимым ресурсом для формирования долгосрочной памяти, а размерность фильтра – ресурс краткосрочной памяти. В то время как в рекуррентной сети ресурс долгосрочной памяти ограничен только особенностями обучения сети, описанных в пункте 1.2 Рекуррентные нейронные сети.

В результате предварительного тестирования с учетом использования всего доступного набора объясняющих переменных для используемых архитектур рекуррентных нейронных сетей было определено временное окно в один квартал. Для сверточной нейронной сети – временное окно в два квартала. Паттерн входных данных в

первом случае представляет собой матрицу объясняющих переменных, с добавленным к ним квадратом целевой переменной на промежутке от t до $t-3$. Выходным обучающим паттерном является значение целевой переменной в $t+1$ период времени. Таким образом, запаздывание всех объясняющих факторов и квадрата целевой переменной равняется одному месяцу. Важно отметить, что при построении прогноза с использованием рекуррентных нейронных сетей использование в числе объясняющих факторов непосредственно целевой переменной с некоторым лагом, приводит к тому, что прогноз имеет запаздывание. Данный вывод был получен в результате предварительного тестирования. Подобный эффект может иметь следующую интерпретацию: корреляция между объясняемой переменной и ее лагом будет всегда больше, чем корреляция объясняемой переменной и любой другой экзогенной переменной, следовательно в процессе минимизации функции ошибки коэффициент перед этим фактором на первом слое в линейной комбинации факторов перед функцией активации, будет определяться относительно высоким. Это означает, что вклад лага объясняемой переменной при прочих равных в вывод сети будет относительно большим вклада других переменных. В экстремальном случае при близком к единице коэффициенте перед лагом объясняемой переменной и одновременно около нулевых коэффициентах остальных объясняющих факторах, прогноз модели будет очень похож на значение самого лага. Таким образом, при наличии лага объясняемой переменной в числе обучающих образцов, градиентный спуск идет по кратчайшему пути и попадает в локальный минимум, где важность лага целевой переменной признается очень высокой, а остальные практически игнорируются. По этой причине в данной работе вместо лага индекса производства исследуемой отрасли используется квадрат лага индекса. Соответственно для сверточной нейронной сети паттерн входных данных представляет собой матрицу экзогенных факторов и квадрата объясняемой переменной на периоде от t до $t-6$, а целевое значение обучающего паттерна – значение нормированного индекса производства в $t+1$ период времени. Для сверточной нейронной сети, описанная выше проблема с лагом объясняемой переменной сохраняется, поскольку ядро свертки также использует линейную комбинацию значений входных факторов.

В данной главе был проведен первичный анализ имеющихся данных, описана процедура предобработки для последующего использования этих данных в процессе обучения моделей. Располагаемые эмпирические данные включают по 18 показателей финансового и реального сектора экономики для каждой отрасли. Практически для каждого фактора из теоретической модели была подобрана соответствующая прокси-переменная.

4 Эмпирическое моделирование, верификация полученных результатов

Прежде чем приступить к моделированию проводится проверка исследуемых индексов на асимметричность, с применением процедуры Сикхела, а также нелинейного теста Цея [37]. Результаты для индексов с исключенным трендом (для выделения линейного тренда использовался фильтр Ходрика-Прескота) приведены в таблице .

Таблица 2 – Проверка индексов производства на нелинейность

| Индекс | Добыч полезных ископаемых | Прочие полезные ископаемые | Нефтепродукт ы, кокс, ядерные материалы | Металлургия |
|--|---------------------------------|----------------------------------|--|-------------|
| Кол-во наблюдений > среднего в уровнях/в разностях | 88/104 | 96/86 | 93/94 | 80/71 |
| Стд отклонение > среднего в уровнях/в разностях | 0.78/0.47 | 3.55/3.22 | 0.73/0.31 | 13.6/8.23 |
| Кол-во наблюдений < среднего в уровнях/в разностях | 89/73 | 81/90 | 84/ 83 | 97/106 |
| Стд. отклонение < среднего в уровнях/в разностях | 0.93/0.77 | 4.03/3.72 | 0.94/0.48 | 9.64.6 |
| Коэффициент асимметрии в уровнях/в разностях | -0.27/- 1.23*** | -0.18/-0.97*** | -0.45/-0.93*** | 0.51/1.74 |
| Тест Тцея в уровнях/в разностях | 0.001 (6.9) | 0.011 (6.46) | 0.001 (12.3) | 0.35 (0.84) |

Приведенные статистические характеристики индексов дают основания предположить наличие асимметрий двух типов, согласно классификации Сикхела: «крутизны» и «глубины», поскольку для трех из четырех индексов количество наблюдений больше среднего, как в уровнях, так и в разностях, превышает количество наблюдений ниже среднего и удалены друг от друга дальше значения, находящиеся ниже среднего. Но при этом статистически значимый отрицательный коэффициент асимметрии (на 5% уровне значимости) наблюдается только для первых трех индексов в разностях⁴, т.е. в среднем наклон на периодах роста более пологий, чем на периодах снижения. Индекс производства металлургической промышленности и вовсе демонстрирует положительную асимметрию обоих типов, т.е. спады в отрасли имеют более затяжной характер, чем периоды роста.

Тест Цея применялся только к первым разностям, поскольку индексы нестационарны в уровнях. В таблице приведены значения вероятностей и в круглых

⁴ Нулевая гипотеза теста – это не отрицательность коэффициента асимметрии.

скобках статистика теста. В разностях трех индексов из четырех тест отвергает нулевую гипотезу об отсутствии нелинейности в пользу альтернативной – ее наличия

4.1 Проверка гипотез исследования, обучение и выбор оптимальной модели

В рамках систем одновременных уравнений, приведенных в разделе 2.3. Определение объясняющих факторов и содержательные гипотезы, или любых других линейных эконометрических моделях, не имеет смысла проверять гипотезу о статистической значимости коэффициентов перед факторами в данном случае, поскольку значимость коэффициента говорит о наличии линейной связи. В нейронной сети, как будет показано далее, прогностические способности фактора могут меняться со временем, как по силе своего влияния на объясняющую переменную, так и по направленности влияния. Следовательно, для проверки сформулированных гипотез необходимо использовать специальные подходы, позволяющие интерпретировать результаты нелинейных моделей машинного обучения. В данном случае будет использоваться сочетание двух подходов: «Shap» [38] (Shapley additive explanations - Аддитивное объяснение Шепли), который основан теории классического вектора Шепли из теории игр, и «DeepLift», о котором далее пойдет речь.

В случае деревьев решений модель поочередно оценивается с интересующим фактором и без него во всевозможных перестановках – очередности подачи факторов в модель. Усредненная разница между ошибками при всех перестановках – это и есть значение вектора Шепли для отдельного фактора. Однако нейронные сети не могут обрабатывать нулевые входные объекты, поэтому возникла необходимость переопределить влияние относительно чего-то другого.

В работе [39] представлен подход к интерпретации выводов глубоких нейронных сетей – «DeepLift». Разработанный авторами подход – это метод декомпозиции прогноза вывода нейронной сети по конкретному входу путем обратного распространения вкладов всех нейронов в сети для каждого входного фактора. Метод сравнивает активацию каждого нейрона с его «эталонной активацией» и присваивает баллы вклада в соответствии с разницей. Метод может также выявить зависимости, которые упускают другие подходы интерпретации выводов моделей, обучаемых методом обратного распространения ошибки, поскольку, принимаются во внимание положительные и отрицательные вклады факторов.

В конечном итоге предметом изучения является то, как выходные данные изменились по сравнению с некоторым базовым значением, когда изменились входные данные по отношению к базовому значению. Для того, чтоб рассчитать подобное, используется концепция интегрированного градиента. Подход, используемый

интегрированными градиентами, заключается в том, чтобы задать следующий вопрос: что можно вычислить, что является аналогией градиента, который также признает наличие базового значения. Интегрированный градиент – это интеграл градиентов между базовым значением и интересующей точкой. На самом деле вычислить интеграл градиентов сложно, поэтому вместо этого он аппроксимируется с помощью суммы Риммана⁵: градиенты берутся с большим количеством маленьких шагов между базовым значением и интересующей точкой. В оригинальной статье эту сумму называют «мультипликатором».

Минимизируя функцию ошибки в процессе работы градиентного спуска мы вычисляем градиенты по параметрам, определяя таким образом направление движения по поверхности ошибки в сторону ее минимизации (если используем отрицательный градиент). В случае расчета важности фактора для каждого слоя необходимо вычислить данный мультипликатор вместо градиента. Отклонение фактора от базового значения, умноженное на этот мультипликатор – это и есть важность фактора. Однако в таком виде данную процедуру уже можно имплементировать в алгоритм обратного распространения, т.е. можно выполнить обратное распространение по этим множителям, чтобы найти угловой коэффициент входных данных по отношению к выходным данным модели, что и позволяет определять важность фактора. В случае регрессии, базовое предположение – это среднее значение по выборке. Относительно него происходит измерение влияния переменных.

В отличие от большинства методов на основе градиента, использование отличия от эталона позволяет «DeepLift» распространять сигнал важности даже в ситуациях, когда градиент равен нулю, и позволяет избежать артефактов, вызванных неоднородностями градиента. Поскольку оценки «DeepLift» вычисляются с использованием алгоритма, подобного обратному распространению, их можно достаточно быстро получить за один проход в обратном направлении после того, как был сделан прогноз.

Существуют и другие подходы [40, 41], которые вносят возмущения в отдельные входы или нейроны и наблюдают за воздействием на более поздние нейроны в сети. Однако такие методы могут быть неэффективными с вычислительной точки зрения, поскольку каждое возмущение требует отдельного прямого распространения по сети. Они также могут недооценивать важность функций, которые превысили свой вклад в результат.

Используемая в работе программная реализация данного подхода из пакета Shar 0.36.0 для языка программирования «Python» отличается от описанной выше за счет использования распределения фоновых выборок вместо одного эталонного значения и

⁵Сумма Римана — один из механизмов определения интеграла через сумму вида $\sum f(x)\Delta x$.

использования уравнений Шепли для линеаризации таких компонентов, как «max», «softmax». Таким образом, для интерпретации полученных результатов используется подход, объединяющий две идеи: «Shap» и «DeepLift». Далее, для краткости, вектор оценок степени влияния фактора на прогнозируемое значение мы будем называть вектором Шепли.

Таким образом, проверка гипотезы относительно влияния отдельного фактора на объясняемую переменную сводится к определению степени важности данного фактора в точности построения нейронной сетью правильного прогноза. При этом в рамках выбранного подхода интерпретации полученных результатов также возможно определить направленность влияния, которую оказывает определенный фактора на целевую переменную при построении прогноза.

Далее приведем результаты прогнозирования для индексов производства трех отраслей с использованием индивидуальных прогнозных моделей с архитектурой сверточной нейронной сети, описанной в главе 2 Разработка модели прогнозирования индексов производства отечественной экономики с учетом межотраслевых связей. Результаты обучения и прогноз на тестовой выборке, индивидуальной модели, для отрасли добыча топливо-энергетических полезных ископаемых приведен на рисунке .

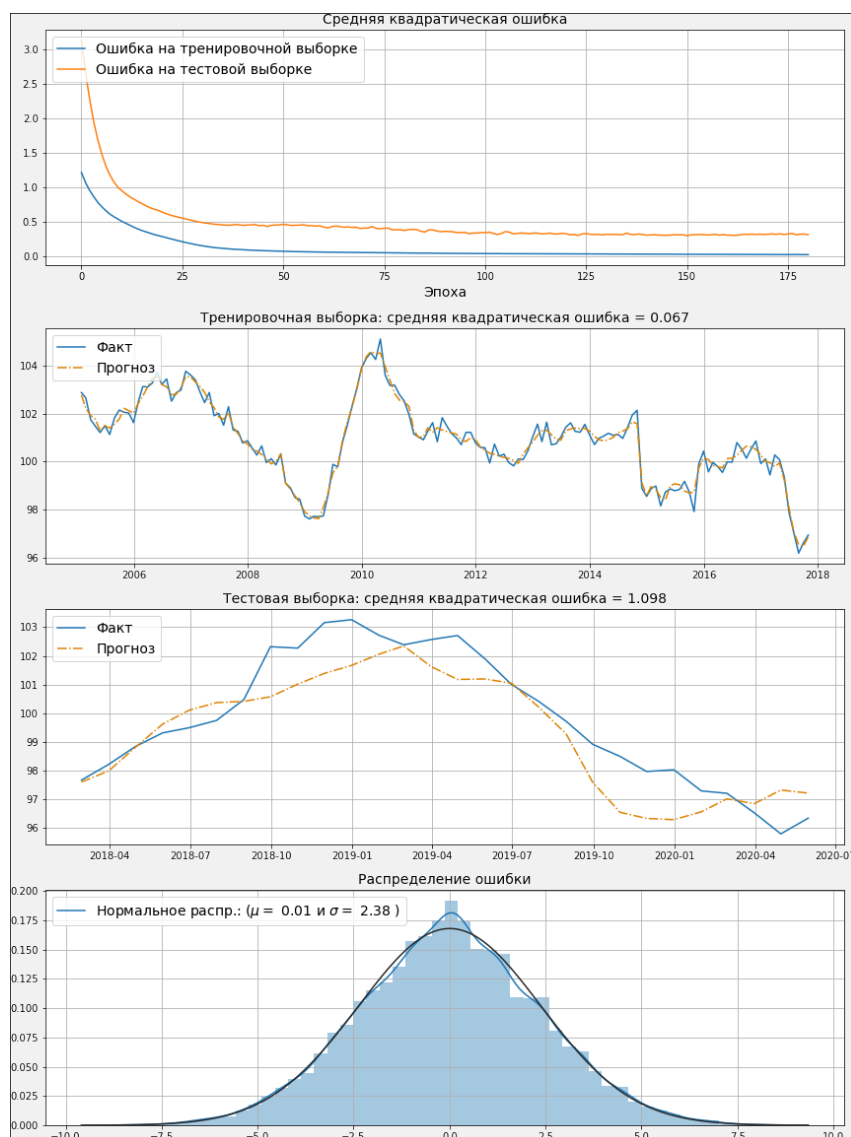


Рисунок 12 – Результат прогнозирования для индекса добычи топливо-энергетических полезных ископаемых, полученный при помощи индивидуальной нейросетевой модели

Первый график рисунка демонстрирует процесс обучения нейронной сети. Оранжевая кривая – это изменение значения функции ошибки с прохождением каждой эпохи. В данном случае обучение остановилось на 177 эпохе. При этом ошибка на тестовой выборке синхронно снижалась на протяжении всего процесса обучения, за исключением последних десяти эпох, где был зафиксирован выход на плато, после чего была произведена автоматическая остановка процесса обучения. Второе изображение на рисунке является восстановленным выводом нейронной сети на обучающей части выборки. Значение ошибки на тренировочной выборке на последней эпохе равняется в данном случае 0,067. Третье изображение – прогноз нейронной сети, приведенный к естественной шкале значений объясняемой переменной с использованием математического ожидания и дисперсии, оцененных по тренировочному набору данных. Последнее обеспечивает естественные условия для прогнозирования, когда целевое значение еще не доступно. В данном случае ошибка сети на тестовой выборке равняется

1.098. В дальнейшем, после полученных прогнозов с использованием других архитектур нейронных сетей, будет проведен тест на различие ошибок между моделями. Последний график на рисунке представляет собой функцию плотности распределения ошибки на обучающей выборке, от куда следует, что ошибка распределена нормально с нулевым математическим ожиданием и среднеквадратическим отклонением 2,38.

Далее обратимся к результатам построения векторов Шепли для объясняющих факторов, что позволит интерпретировать результаты обучения нейронной сети. На рисунке представлено разложение прогноза одной точки по вкладу в него всех факторов с задействованием только одного лага (предыдущего дня). Выходное значение – это предсказание модели для конкретного наблюдения. В данном случае было использовано первый паттерн из тестовой выборки.

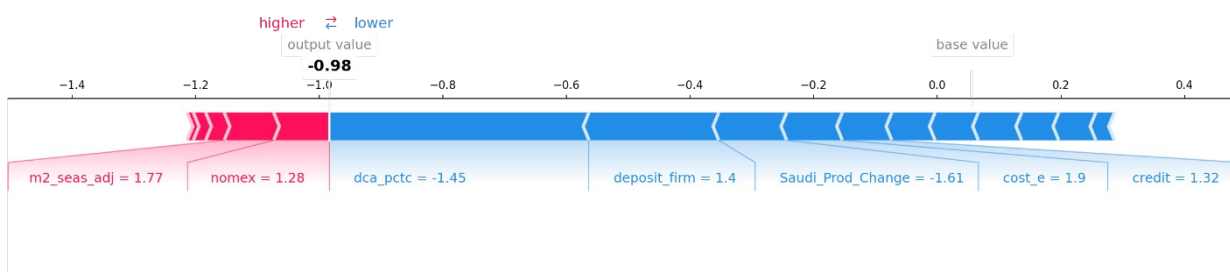


Рисунок 13 – Разложение прогноза одной точки для индекса добычи топливно-энергетических полезных ископаемых

На рисунке приведены результаты построения векторов Шепли по описанной выше методологии для каждой прогнозной точки индекса производства топливно-энергетических полезных ископаемых на тестовой части выборки. Факторы на графике упорядочены в соответствии с общей степенью их влияния на формирование целевой переменной – прогнозируемого индекса производства. Исходя из графика можно сказать следующее: лаговое значение квадрата индекса производства, депозиты фирм, обменный курс рубля и кредиты, выданные некоммерческим организациям, оказывают самое значительное влияние формирование прогноза. При этом низкие значения фактора, как правило связаны с низкими значениями целевой переменной и наоборот, но при этом низкие значения оказывают более значимое влияние на формирование прогноза в сторону его понижения. Так же можно наблюдать некоторый порог, с которого относительно высокие значения лага (выше среднего) начинают оказывать отрицательное влияние на формирование прогнозного значения. Депозиты фирм имеют некоторое нелинейное влияние: для некоторых прогнозных значений относительно низкие объемы размещенных депозитов организаций оказывали слабо отрицательное влияние на индекс производства, но при этом высокие положительные значения оказывали строго отрицательное влияние. Влияние обменного курса также нелинейно: высокое значение курса рубля в большинстве

случаев ассоциировалось нейронной сетью с высоким индексом производства. Это говорит о том, что в периоды ослабления национальной валюты фирмы, занимающиеся добычей полезных ископаемых, наращивают объемы производства для того, чтоб получать сверхприбыль в рублевом эквиваленте за счет экспорта. При этом низкие значения обменного курса рубля так же оказывают положительное влияние на индекс производства, но уже слабее по степени оказываемого влияния. Экономическая интерпретация может быть следующая: отрасль за счет ослабления рубля использует рост внутреннего спроса на свою продукцию, который возникает в результате замещения относительно подорожавших импортной продукции.

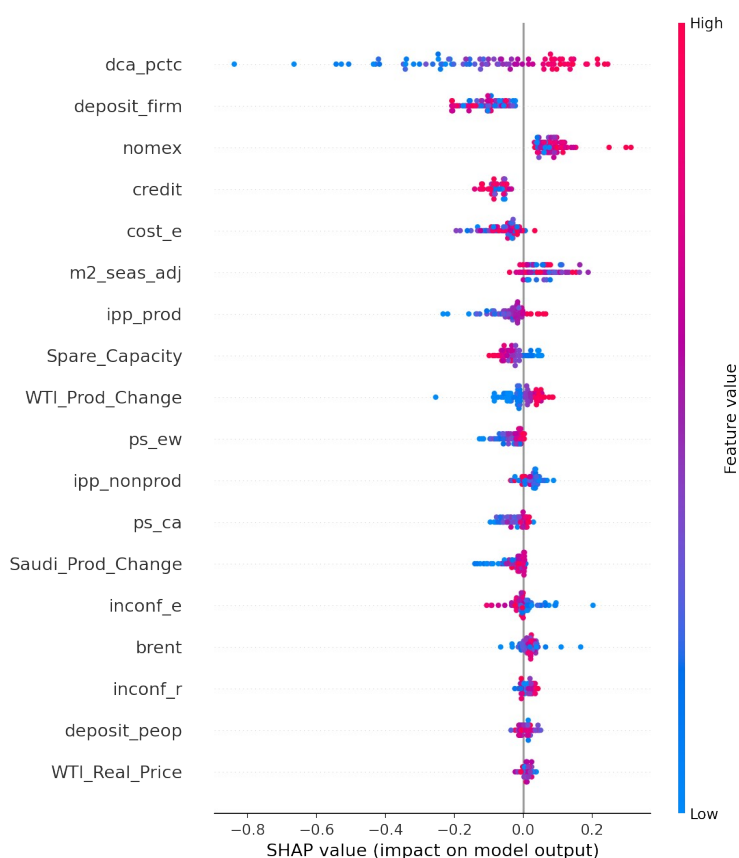


Рисунок 14 – Анализ направленности влияния факторов при построении прогноза для индекса добычи топливно-энергетических полезных ископаемых

На рисунке приведены усредненные значения Шепли для всех объясняющих переменных прогнозной модели для индекса топливно-энергетических полезных ископаемых. В отличие от предыдущей диаграммы. Где значение Шепли было приведено для каждой отдельной прогнозной точки, в данном случае мы имеем усредненное по модулю значение по всем наблюдениям тестовой выборки. Это позволяет понять общую силу влияния, которую имеет фактор на прогнозируемую величину.

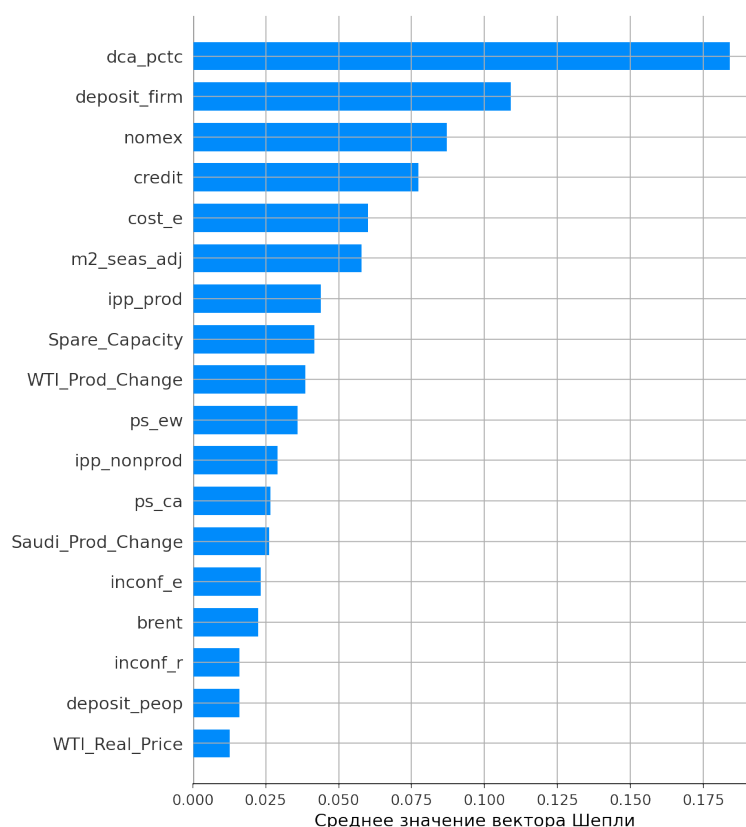


Рисунок 15 – анализ степени влияния объясняющих факторов на прогнозируемое значение индекса добычи топливно-энергетических полезных ископаемых

Далее перейдем к результатам построения прогнозов для отрасли производства кокса, нефтепродуктов и ядерных материалов.

На рисунке приведены результаты обучения и прогнозирования на тестовой части выборки нейронной сетью. Первое изображение демонстрирует непосредственно процесс обучения. Из него следует, что ошибка на тренировочной выборке снижалась постоянно на все периоде обучения. Ошибка на тренировочной выборке вышла практически на плато уже на 20 эпохе обучения. Но свидетельств переобучения нет, поскольку в дальнейшем не происходило резкого роста ошибки на тестовой выборке и автоматическая остановка обучения на 65 шаге обучения выглядит разумно. Второй график на изображении представляет собой восстановленный прогноз на тренировочной выборке. Прогнозная кривая также не демонстрирует признаков переобучения, поскольку не повторяет в точности фактические данные, следовательно, можно ожидать, что сеть не потеряла обобщающую способность. Восстановленный прогноз в естественном диапазоне значений индекса производства приведен на третьем графике. Средняя квадратическая ошибка увеличивается приблизительно в два раза при переходе от обучающей к тестовой выборке. При этом остатки модели на тренировочной выборке имеют распределение близкое к нормальному, что говорит о том, что у модели нет склонности совершать

систематическую ошибку по переоценке или недооценке по отношению к истинному значению прогнозируемой переменной.

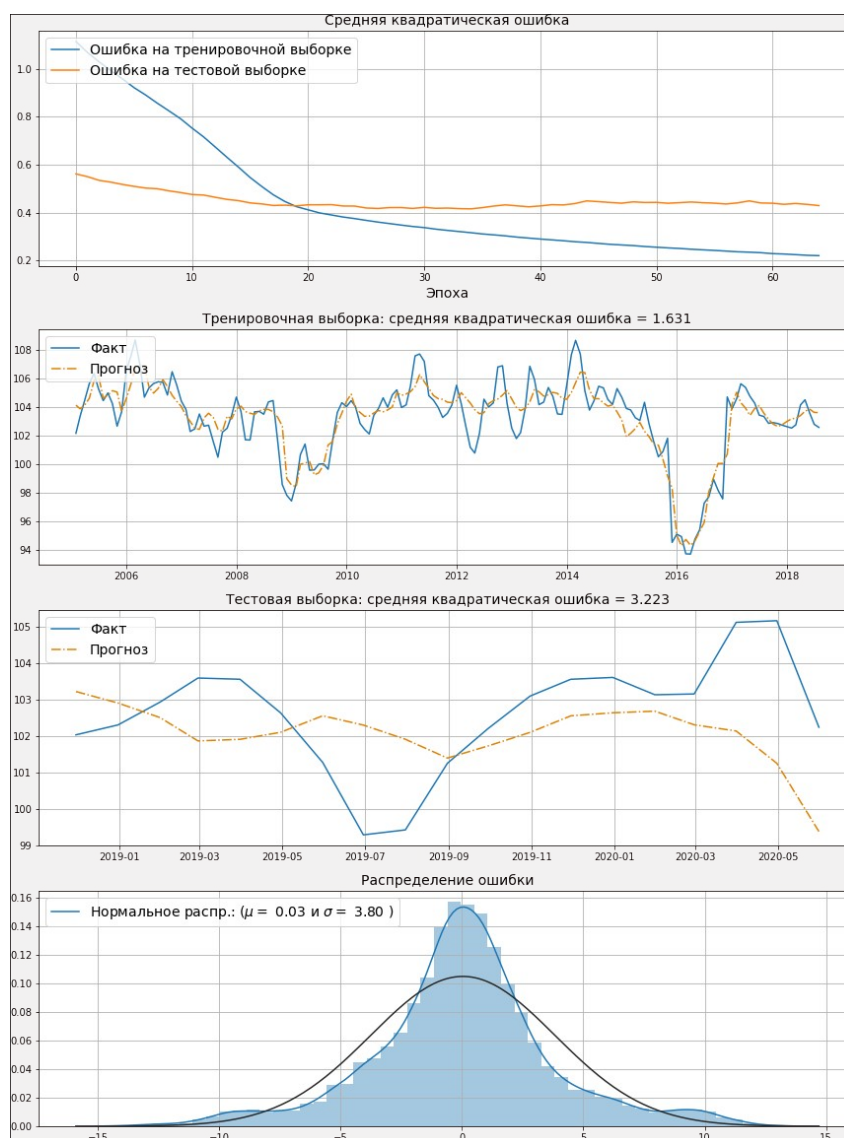


Рисунок 16 – Результат прогнозирования для индекса производства кокса, нефтепродуктов и ядерных материалов при помощи индивидуальной нейросетевой модели

На рисунке приведено разложение прогноза первой тестовой точки в данных по факторам влияния. Ширина диапазона блока, соответствующего определенному фактору, говорит о степени влияния на целевую переменную. Из разложения следует, что курс рубля оказал наибольшее по модулю влияние на конкретное спрогнозированное значение индекса. При этом укрепление курса рубля привело к снижению индекса производства. Это говорит о том, что в данном случае производители отрасли в большинстве своем столкнулись со снижением спроса со стороны зарубежных рынков в ответ на относительное удорожание их продукции за рубежом, в ответ на что вынуждены были снизить объемы производства.

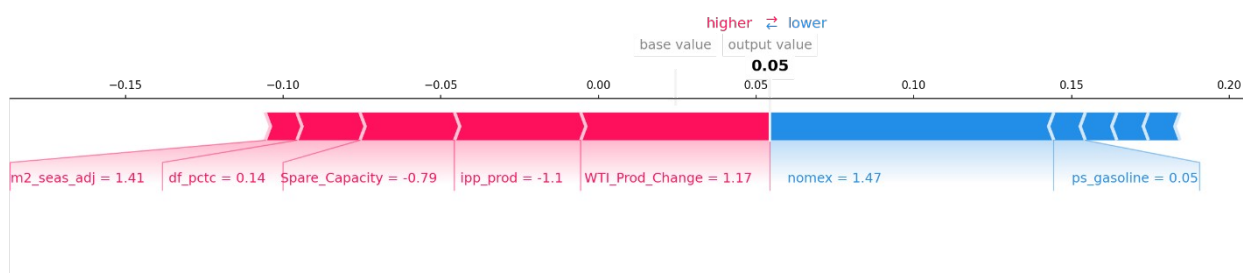


Рисунок 17 – Анализ направленности влияния факторов при построении прогноза для производства кокса, нефтепродуктов и ядерных материалов

На рисунке приведен анализ влияния факторов на формирование прогнозных значений на тестовой выборке. Согласно анализу, низкие значения средней стоимости нефти различных марок оказывают значительно отрицательное влияние на индекс производства отрасли. Высокие средние цены на нефть слабо-положительно связаны с индексом производства. Это может иметь следующее объяснение: низкая стоимость нефти снижает спрос на альтернативные нефти энергоносители: уголь и ядерное топливо со стороны внешних рынков. В ответ на что производители снижают предложение. К тому же, отрицательно связанный со стоимостью нефти курс, ослабевая по отношению европейским валютам, оказывает отрицательное влияние на покупательскую способность населения, которое сокращает спрос на продукцию отрасли, вследствие чего производители вынуждены еще в большей степени сокращать предложение. Низкая продовольственная инфляция оказывает положительное влияние на индекс производства отрасли. Высокая продовольственная инфляция имеет менее выраженное влияние, но уже отрицательное. Это говорит о том, что население, обладая высокой покупательской способностью, создает высокий спрос на продукцию отрасли и стимулирует производителей производить больше и, соответственно, имея низкую покупательную способность, население в относительно меньшей степени сокращает спрос и вынуждает производителей снижать объемы производства. Другими словами, спрос населения является слабо эластичным при снижении покупательской способности и более эластичным при улучшении покупательской способности.



Рисунок 18 – Анализ направленности влияния факторов при построении прогноза для индекса производства кокса, нефтепродуктов и ядерных материалов

На рисунке приведен ранжированный по убыванию степени влияния список факторов. Суммарно по модулю наибольшее влияние на формирование прогнозного значения индекса производства кокса, нефтепродуктов и ядерных материалов оказывала средняя цена нефти марок Brent, Dubai и WTI. На втором месте находится квадрат лага объясняемой переменной. Далее по убыванию степени влияния идут: реальная стоимость нефти марки WTI, кредиты, выданные некоммерческим организациям, резервные запасы нефти стран ОПЕК, продовольственная инфляция, изменение объемов производства тexasской нефти и т.д.

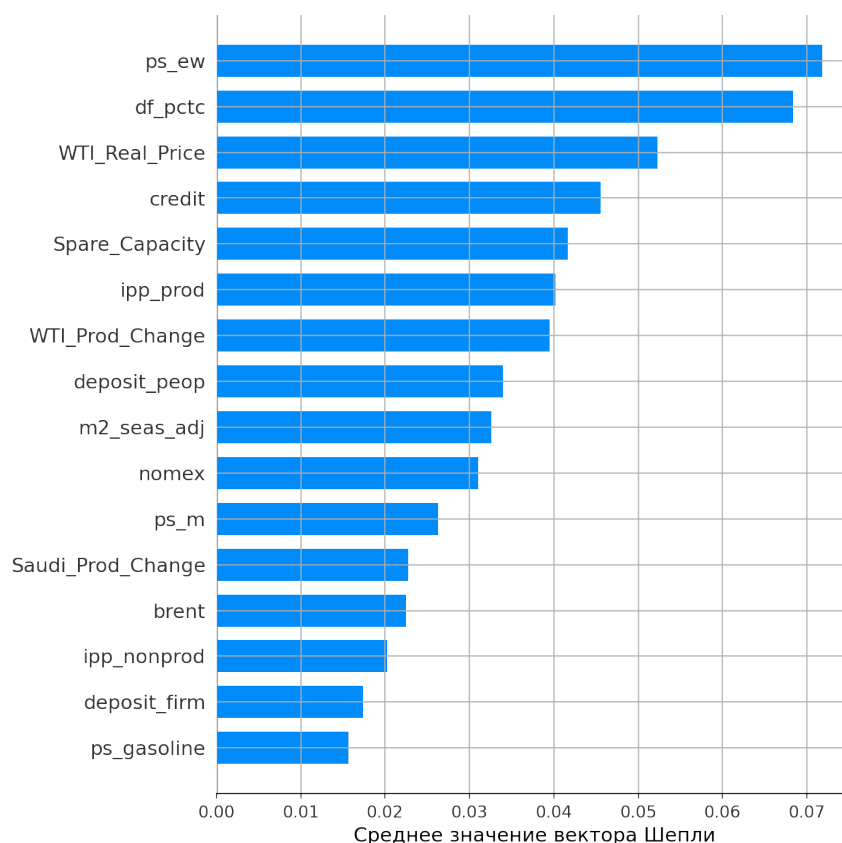


Рисунок 19 – анализ степени влияния объясняющих факторов на прогнозируемое значение индекса производства кокса, нефтепродуктов и ядерных материалов

Далее будет построен прогноз для тестовой выборки индекс производства химическая отрасли. На рисунке приведен процесс обучения нейронной сети и результаты применения модели к тестовой части выборки. Процесс обучения характеризуется постоянным снижением ошибки прогноза, как на обучающей, так и на тестовой выборках на протяжении 200 эпох обучения. На втором изображении рисунка приведен восстановленный прогноз на тренировочной выборке.

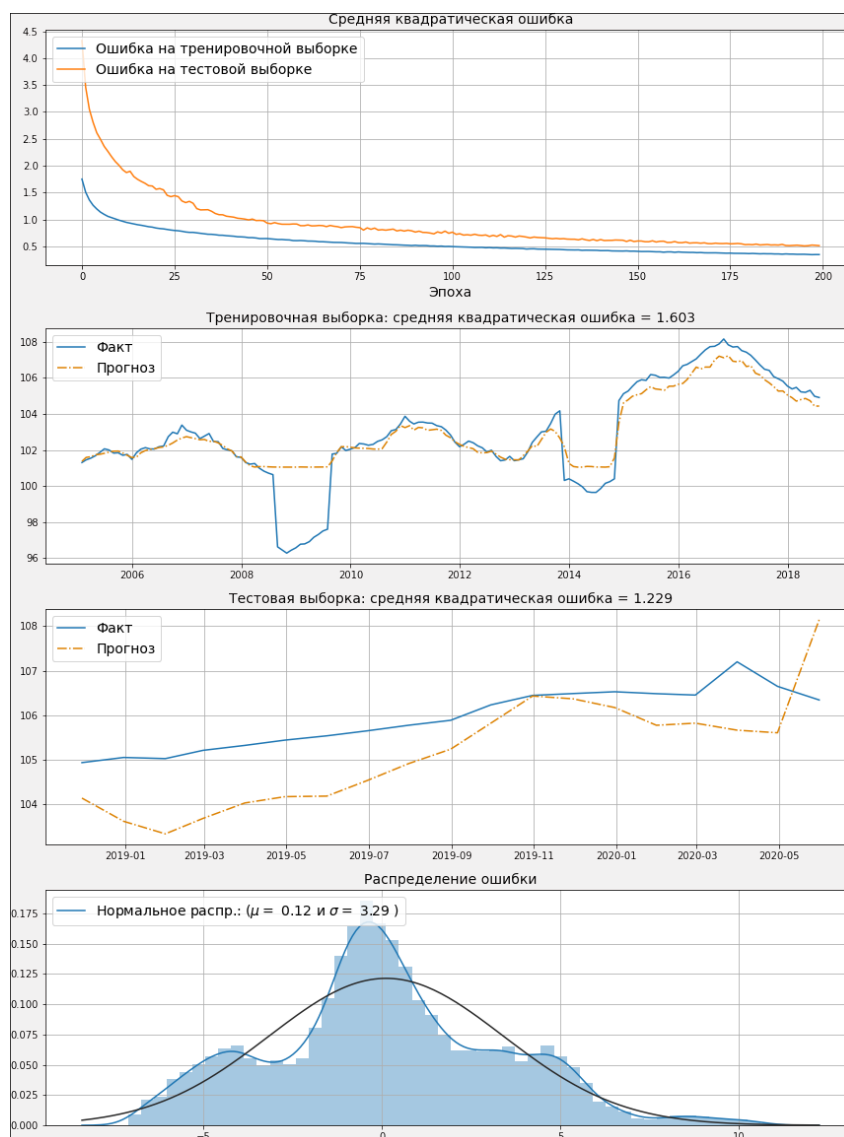


Рисунок 20 – Результат прогнозирования для индекса производства химической отрасли, построенный при помощи индивидуальной нейросетевой модели

На третьем графике рисунка приведен восстановленный прогноз в истинных масштабах определения значений индекса производства. Среднеквадратическая ошибка на тестовой части составляет 1,22 (на обучающей 1,603). Незначительное отличие ошибки на тестовой выборке от обучающей говорит о том, что сеть в процессе обучения не потеряла обобщающую способность, т.е. не переобучилась. Распределение ошибки прогноза при этом значительно отличается от нормального и говорит о том, что имеется тенденция к накоплению ошибок в плюс и минус пять стандартных отклонений от среднего.

На рисунке приведено разложение первой прогнозной точки по факторам, оказавшим на нее формирование для первого лага. Всего, в данном случае, сеть использует три лага каждой объясняемой переменной, чтобы построить прогноз. Так как все факторы нормированы, то среднее каждого фактора равно нулю. Это в некоторой степени упрощает интерпретацию диаграммы. Исходя из полученных результатов можно

сказать, что положительное отклонение рубля от его нулевого среднего оказывает понижающее влияние на индекс производства отрасли. Это говорит о том, что производители химической отрасли вынуждены сокращать объемы производства, когда позиции рубля по отношению к доллару ослабевают. Возможно, это связано с тем, что факторы производства закупаются за рубежом и, следовательно, удешевление отечественной валюты приводит к росту издержек производства.

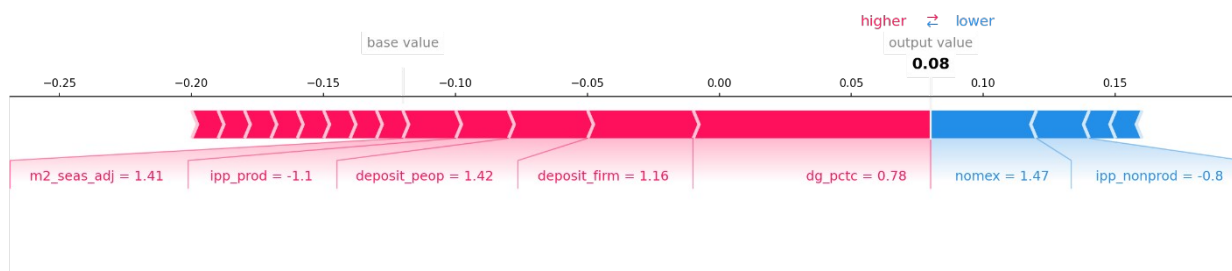


Рисунок 21 – Анализ направленности влияния факторов при построении прогноза для производства химической отрасли

Индекс непроизводственной инфляции в данном случае имеет отрицательное отклонение от среднего (-0,8). Понижение продовольственной инфляции связано с более низкими значениями индекса производства, что имеет свою экономическую интерпретацию. Непродовольственная инфляция может отражать рост спроса на непродовольственную продукцию (чем и является продукция, производимая отраслью), а следовательно, снижение непроизводственной инфляции может быть связано со снижением спроса. В ответ на снижение спроса производители сокращают производство. Снижение продовольственной инфляции напротив оказывает положительное влияние на индекс производства отрасли. Уровень продовольственной инфляции снижается в периоды относительного спокойствия в экономике, когда население, как один из потребителей продукции отрасли, имеет высокую покупательную способность и, следовательно, создает дополнительный спрос на продукцию химической отрасли. В ответ на это производители повышают объем производства. При этом, чем выше изменения денежной массы, тем выше индекс производства. Однако денежная масса имеет отрицательную корреляцию с продовольственной инфляцией, что объясняет выявленную связь, хотя она и противоречит макроэкономической теории. Квадрат лага целевой переменной в данном случае положительно связан с индексом и имеет наибольшее по модулю влияние. Депозиты фирм и населения имеют в данной точке положительное отклонение от среднего и оказывают положительное влияние на формирование прогнозного значения индекса.

На рисунке приведено влияние факторов на формирование каждого прогнозируемого уровня индекса производства. Динамика рубля имеет нелинейное

влияние на индекс производства химической отрасли: как высокие, так и низкие значения курса оказывали отрицательное влияние на формирование прогнозируемого значения. Возможно, это говорит о том, что в действительности есть некоторая предельно убывающая полезность от укрепления рубля для производителей отрасли. При крайне слабом курсе производители не могут закупать импортные факторы производства и средства производства, что заставляет их снижать предложение, а при крайне сильном курсе снижается спрос на продукцию отрасли, как со стороны внешних рынков, поскольку продукция становится относительно дороже, так и со стороны внутреннего рынка, поскольку потребители переключаются на потребление зарубежной продукции, поскольку она становится доступнее.



Рисунок 22 – Анализ направленности влияния факторов при построении прогноза для индекса производства химической отрасли

Депозиты фирм при построении прогнозных значений нейронной сетью, как правило, имели положительное влияние на выпуск. При этом высокие положительные значения чаще ассоциировались с высоким индексом производства. Низкие значения, т.е. отрицательные отклонения от среднего оказывали незначительное положительное влияние на индекс производства. Депозиты людей напротив: низкие значения ассоциировались нейронной сетью с высоким уровнем производства. При этом при

подобных паттернах наблюдалось относительно высокое влияние показателя на вывод модели.

На рисунке представлено относительное среднее влияние всех факторов по модулю. Квадрат лага фактора, как и в предыдущих двух отраслях, оказывает в целом наибольшее влияние на формирование прогноза индекса производства химической отрасли. Далее, с более, чем трехкратным снижением степени влияния находятся обменный курс рубля к доллару США, за которым следуют: депозиты фирм, издержки на электроэнергию, кредиты, выданные нефинансовым организациям, резервные запасы нефти стран ОПЕК и так далее.

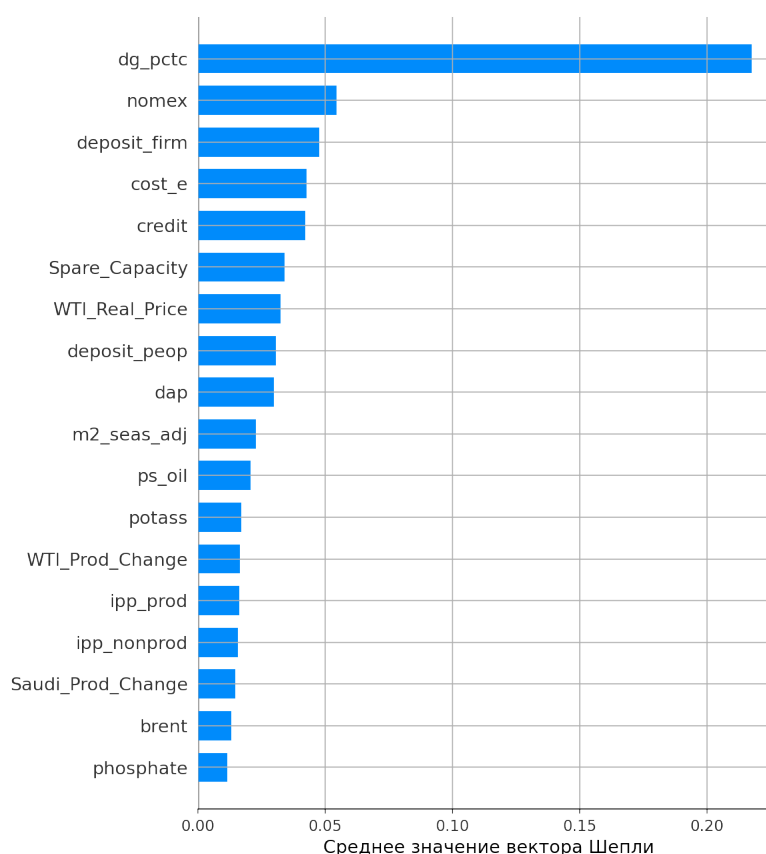


Рисунок 23 – анализ степени влияния объясняющих факторов на прогнозируемое значение индекса производства химической отрасли

В таблице приведены среднеквадратические ошибки на тестовой выборке для рекуррентных и сверточных индивидуальных прогнозных моделей для рассматриваемых индексов производства.

Таблица 3 – среднеквадратические ошибки индивидуальных моделей для прогнозирования индексов производства отраслей

| | Индекс добычи | Индекс производства нефтепродуктов | Индекс химической отрасли |
|-------------------|---------------|------------------------------------|---------------------------|
| Рекуррентная сеть | 1.098 | 3.223 | 1.229 |
| Сверточная сеть | 1.152 | 3.208 | 1.543 |

Из сравнительных результатов следует, что для индекса добычи топливно-энергетических полезных ископаемых наилучший вневыборочный прогноз с точки зрения минимальной среднеквадратической ошибки удалось построить рекуррентной нейронной сети. Для индекса производства кокса, нефтепродуктов и ядерных материалов лучшей оказалась сверточная нейронная сеть, не смотря на менее гладкую кривую ошибки процесса обучения на тестовой выборке.

Далее построим стек нейросетевых моделей и сравним результаты с лучшими моделями из индивидуальных нейронных сетей. На рисунке приведены результаты обучения модели стека и показан процесс изменения индивидуальной ошибки и восстановленный прогноз для обучающей и тестовой выборок для индекса добычи топливно-энергетических полезных ископаемых.

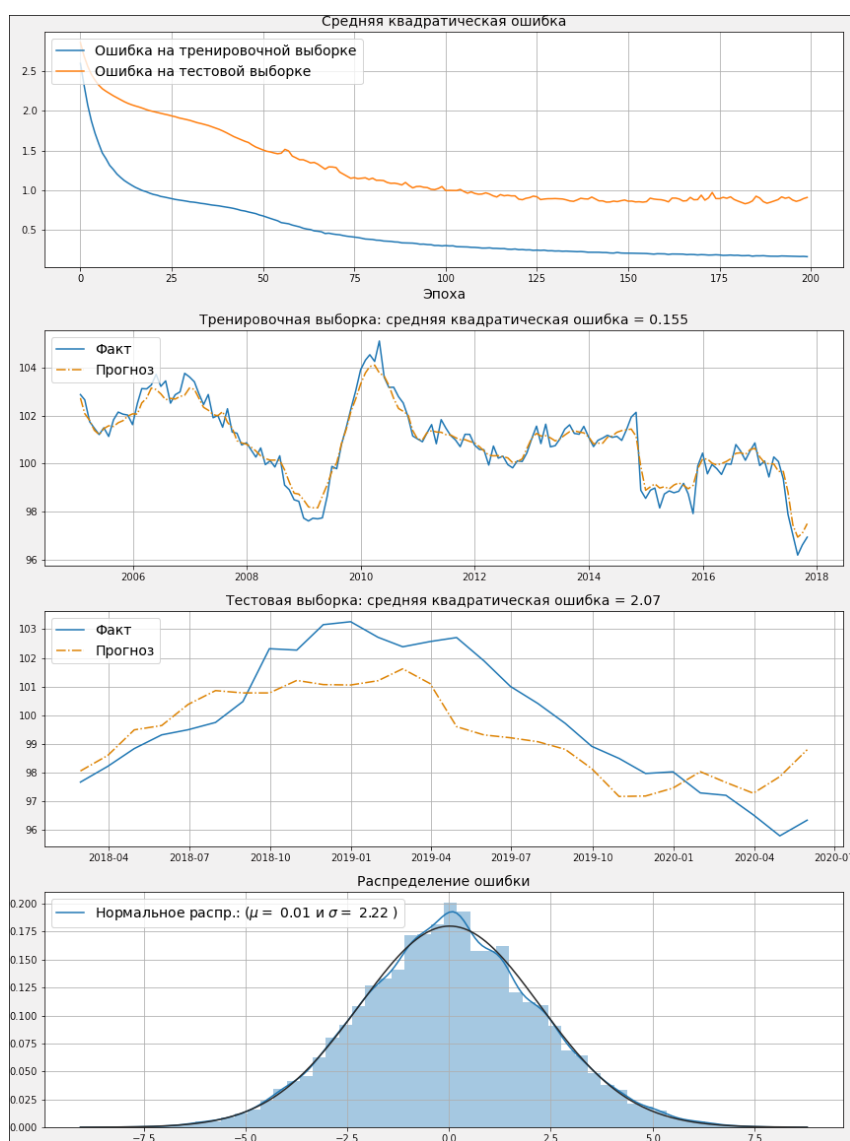


Рисунок 24 – Результат прогнозирования для индекса добычи топливно-энергетических полезных ископаемых при помощи стека нейросетевых моделей

На рисунке приведены кривые ошибок и результаты прогноза той же самой модели стека, но уже для индекса производство кокса, нефтепродуктов и ядерных материалов.

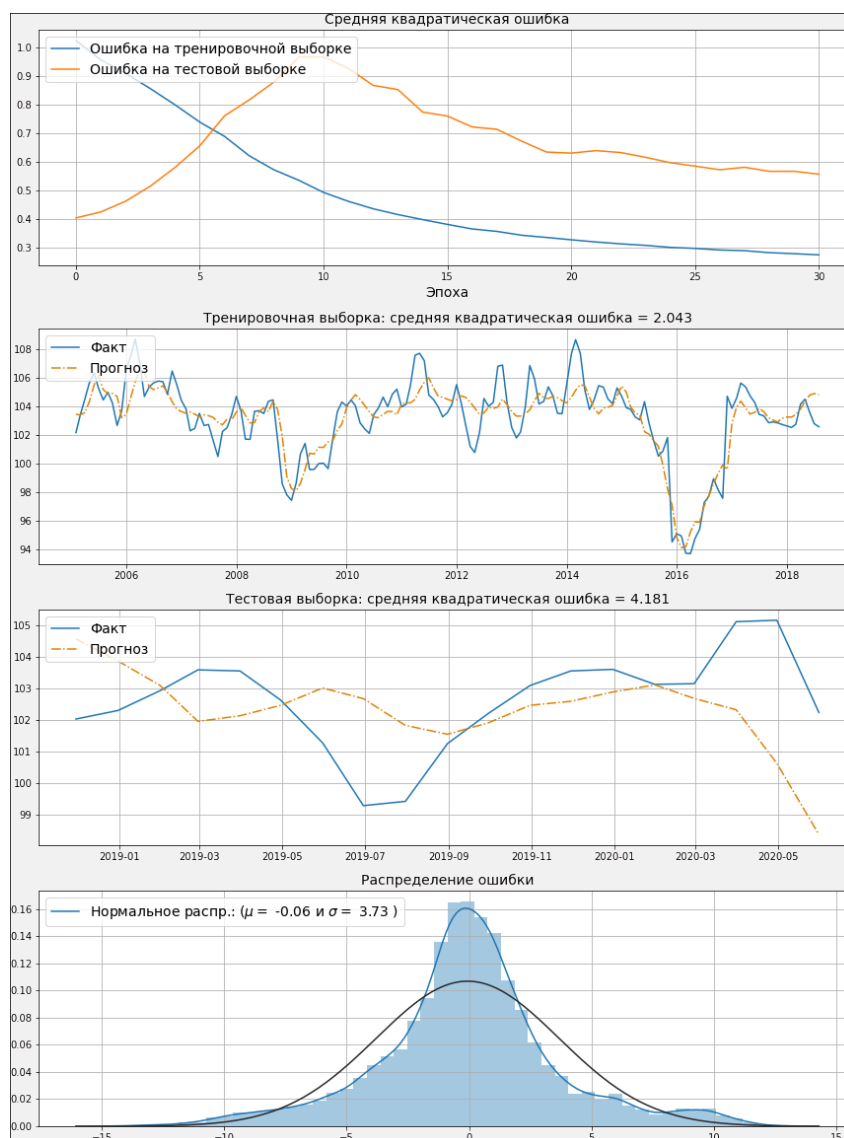


Рисунок 25 – Результат прогнозирования для производства кокса, нефтепродуктов и ядерных материалов при помощи стека нейросетевых моделей

На рисунке приведены результаты прогноза и процесс обучения модели стека для индекса производства химической отрасли.

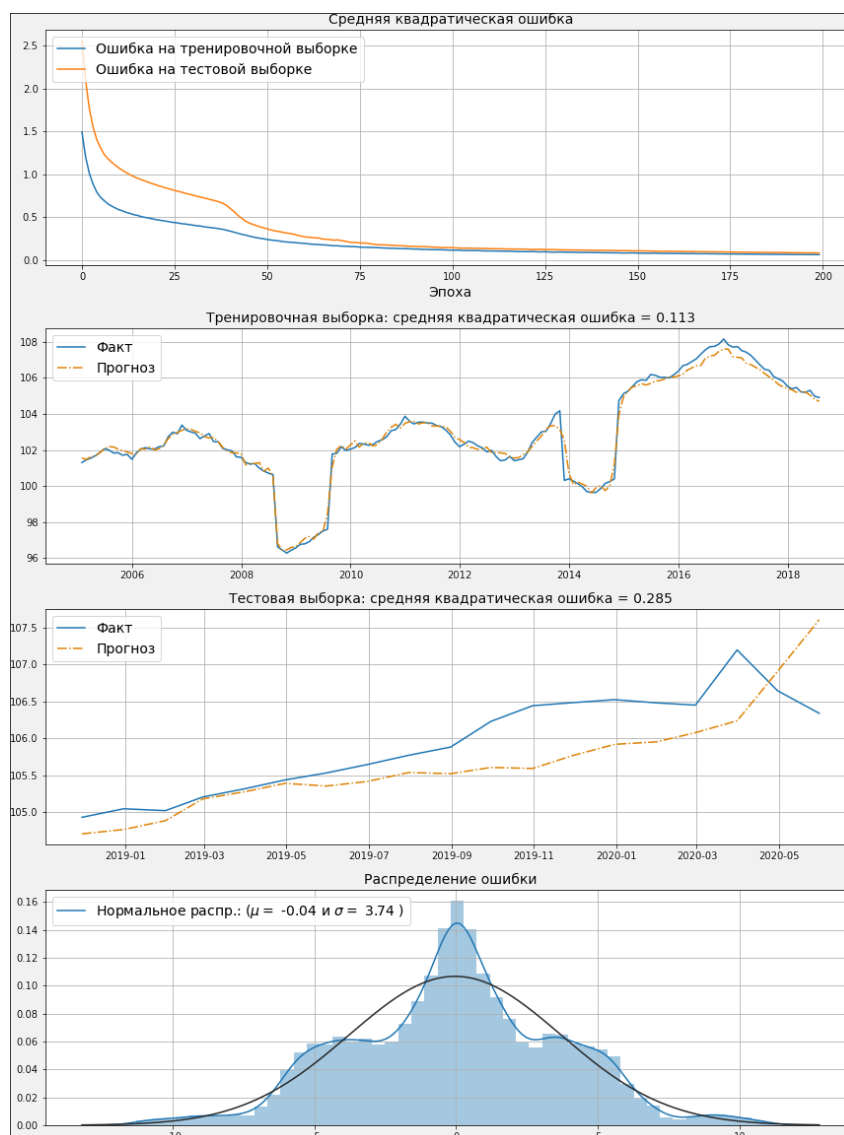


Рисунок 26 – Результат прогнозирования для производства химической отрасли при помощи стека нейросетевых моделей

В целом по индивидуальным остаткам можно сказать, что модель стека строит сопоставимые по точности прогнозы по сравнению с индивидуальными моделями. В таблице приведены результаты сравнения прогностических качеств модели стека с наилучшей индивидуальной моделью, также прогноз ARIMA⁶ для каждого изучаемого индекса производства с применением статистического теста Диболда-Мариано.

⁶ Параметры ARIMA для индекса каждой отрасли были определены на основе минимального значения информационного критерия AIC. Прогноз строился при прочих равных: индексы были нормированы, параметры модели оценивались только на тестовой части выборки.

Таблица 4 – Сравнение прогностических качеств безусловных и условных моделей

| Анализируемые отрасли | Сравнение ARIMA и наилучшей модели НС | | Сравнение индивидуальной НС и стека НС | |
|---|---------------------------------------|------------|--|------------|
| | Статистика DM | P-значение | Статистика DM | P-значение |
| Добыча топливно-энергетических полезных ископаемых | 1.997 | 0.046 | -4.073 | 0.001 |
| Производство кокса, нефтепродуктов и ядерных материалов | 1.189 | 0.241 | -1,214 | 0,261 |
| Химическое производство | -2.845 | 0.008 | 2.235 | 0.034 |

Нулевая гипотеза теста состоит в том, что различия между прогнозными свойствами двух моделей отсутствуют, т.е. разность среднеквадратических ошибок для двух моделей статистически неотличима от нуля. Тест проводился для ошибок, рассчитанных на тестовой выборке. Для индекса добычи топливно-энергетических ископаемых гипотеза об отсутствии различий между моделями отвергается, положительная статистика теста показывает, что ошибка индивидуальной модели больше ошибки модели стека, следовательно, модель 2 предпочтительнее модели 1. Для индекса производства кокса, нефтепродуктов и ядерных материалов гипотеза об отсутствии различий между моделями не отвергается, следовательно, модели являются равнозначными. Для индекса химической отрасли гипотеза о равенстве квадратов ошибок отвергается, отрицательная статистика теста показывает, что ошибка модели стека больше ошибки индивидуальной модели, следовательно, индивидуальная нейронная сеть предпочтительнее стека. В части сравнения результатов с бенчмарком, результат также не однозначен. Для отрасли добычи топливно-энергетических полезных ископаемых нейронная сеть продемонстрировала более качественный прогноз, чем бенчмарк ARIMA. Для отрасли производства кокса, нефтепродуктов и ядерных материалов разница между ошибками прогнозов, построенных при помощи двух моделей: нейронной сети и ARIMA статистически неотличима от нуля. Для химической отрасли ошибка нейронной сети выше, чем ошибка бенчмарка. Однако стоит заметить, что прогноз с использованием ARIMA имеет явное запаздывание. Высокая корреляция между значениями t и $t+1$ в модели Бокса-Дженкинса и приводит к тому, что ошибка сопоставима с ошибкой нейронной сети, хотя для последней значительно меньшей степени свойственна проблема запаздывания прогноза. Во много это связано с тем, что среди объясняющих факторов не содержится непосредственно лага самой объясняемой переменной, а используется ее квадрат.

В данной части работы были приведены основные результаты прогнозирования при помощи нейросетевых моделей изучаемых индексов промышленного производства. Использованный подход к интерпретации результатов машинного обучения позволил проверить выдвинутые предположения относительно используемых объясняющих факторов в моделях. В используемом методе степень влияния фактора – это относительная метрика. В основе ее расчета находится следующая идея: мы хотим измерить, на сколько отклониться целевое значение от своего среднего при некотором отклонении фактора от среднего по выборке. Отклонение фактора от среднего по выборке – это и есть значение фактора в t момент времени, когда строится прогноз на $t+1$ момент времени. Отклонение целевой переменной – это оцененное значение прогноза для $t+1$ шага, а его среднее значение – это среднее, по всем прогнозам, обучающей выборки. Для вычисления такой величины используется концепция интегрированного градиента. Вместо частной производной в точке необходимо посчитать интеграл под кривой производной между двумя точками: средним значением фактора и его отклонением в момент времени t . Поскольку значения фактора и целевой переменной отклоняются, как положительно, так и отрицательно от своего среднего, то данный подход позволяет определить направленность влияния фактора на целевую переменную. Применение данного метода позволило выявить ряд нелинейных связей между факторами и целевой переменной. К примеру, между средними ценами на нефть различных марок и индексом производства однозначно положительная связь, но в периоды значительного падения цен фактор имеет относительно большее влияние на прогнозируемое значение. В данном случае зависимость однонаправленная, хотя и нелинейна по степени влияния. Аналогично для резервных запасов нефти стран ОПЕК, но с обратным влиянием. Из анализа влияния обменного курса рубля на индекс производства химической отрасли следует: как высокие, так и низкие значения курса оказывали отрицательное влияние на формирование прогнозируемого значения. Возможно, это говорит о том, что существует некоторая предельно убывающая полезность от укрепления рубля для производителей отрасли. При крайне слабом курсе производители не могут закупать импортные факторы производства и средства производства, что заставляет их снижать предложение, а при крайне сильном курсе снижается спрос на продукцию отрасли, как со стороны внешних рынков, поскольку продукция становится относительно дороже, так и со стороны внутреннего рынка, поскольку потребители переключаются на потребление зарубежной продукции, поскольку она становится доступнее. В части выявленного взаимодействия остальных переменных полученные результаты преимущественно совпали с первоначальными ожиданиями.

ЗАКЛЮЧЕНИЕ

В данной работе были изучены возможности использования сверточных и рекуррентных нейронных сетей для прогнозирования индексов промышленного производства по структуре ОКВЭД: «Добыча топливно-энергетических полезных ископаемых»; «Производство кокса, нефтепродуктов и ядерных материалов»; «Химическое производство». Поскольку индексы имеют асимметрию в периодах роста и падения, нами была выдвинута гипотеза о том, что нелинейные методы позволят улучшить качество прогноза относительно линейных и в практической части исследования мы получаем свидетельства ее неотвержения. Кроме индивидуальных нейросетевых моделей была также разработана и протестирована система нейросетей, объединяющая процесс прогнозирования во взаимосвязанных отраслях. Важное внимание в работе уделяется методам, которые позволяют интерпретировать результаты, получаемые нейросетевыми моделями. Применяемые подходы⁷ позволили провести декомпозицию факторов по степени влияния на вывод сети, а также определить направленность влияния.

При разработке собственной прогнозной модели учитывалась связь между отраслями. К примеру, на стадии добычи топливно-энергетических полезных ископаемых производится сырье для отрасли, производящей кокс, нефтепродукты и ядерные материалы. В свою очередь химическая промышленность использует в качестве сырья в том числе, как продукты переработки второй отрасли, так и первой.

Таким образом, выпуск отрасли добычи топливно-энергетических полезных ископаемых является сырьевым фактором отрасли производства кокса, нефтепродуктов и ядерных материалов, а также химической отрасли. Следовательно, структурная эконометрическая модель выпуска отрасли может включать выпуск другой отрасли, как экзогенный фактор. Однако оценить такую систему уравнений эконометрически и получить из нее прогноз крайне сложно, поскольку в каждом уравнении присутствует проблема эндогенности, к тому же остается нерешенной проблема нелинейности.

Предлагаемая система нейронных сетей позволяет разрешить данные проблемы: прогноз строится одновременно для трех отраслей таким образом, что прогноз одной отрасли влияет на формирование прогноза в другой. Сама прогнозная модель представляет собой комбинацию индивидуальных моделей, каждая из которых может быть использована как самостоятельная. Индивидуальная модель в системе – это

⁷ Аддитивное объяснение Шепли (Shap) [38]; метод на основе интегрированных градиентов (DeepLift) [39].

сверточная нейронная сеть. Ее целевая переменная – индекс производства отрасли, а объясняющие переменные – это факторы спроса и предложения данной отрасли.

Для определения перечня объясняющих факторов и формулирования соответствующих им гипотез для каждой из исследуемых отраслей были построены теоретические системы уравнений «спрос-предложение». К примеру, факторами со стороны спроса для отрасли добычи топливно-энергетических полезных ископаемых являются: биржевые цены нефти, газа, угля; объем производства теплоэнергии; доходы основных потребителей отрасли (химической отрасли, производителей нефтепродуктов); обменный курс рубля и другие. Факторы со стороны предложения: биржевые цены производимых товаров; инвестиции в отрасль; уровень заработной платы; издержки транспортировки; ожидания производителей и другие. Располагаемые эмпирические данные включают по 18 показателей финансового и реального сектора экономики для каждой отрасли. Практически для каждого фактора из теоретической модели была подобрана соответствующая прокси-переменная.

Прежде, чем приступить к построению моделей, в эмпирической части работы были получены свидетельства⁸ асимметричного поведения изучаемых индексов производства. В процессе построения индивидуальных нейросетевых моделей была проведена декомпозиция факторов по степени их влияния на прогноз. Полученные степени влияния – это относительные величины, которые показывают, как отклоняется целевая переменная от своего базового значения (среднее по восстановленному прогнозу) в ответ на отклонение фактора от своего среднего значения. К примеру, наибольшее влияние на формирование прогнозируемых значений индекса производства кокса, нефтепродуктов и ядерных материалов оказывала средняя цена нефти марок Brent, Dubai и WTI. На втором месте находится квадрат лага объясняемой переменной. Далее по убыванию степени влияния: лаг инвестиций, резервные запасы нефти стран ОПЕК, непродовольственная инфляция, изменение объемов производства техасской нефти и другие. Кроме того, используемый метод интерпретации результатов позволил получить не только оценку вклада переменной, но и определить направленность влияния. Оценка влияния производится для каждого прогнозируемого значения отдельно, в результате чего можно сделать вывод о наличии нелинейных зависимостей. К примеру, средними ценами на нефть различных марок и индексом производства однозначно положительная связь, но в периоды значительного падения цен фактор имеет относительно большее влияние на прогнозируемое значение. В данном случае зависимость однонаправленная, хотя и

⁸ Для трех индексов в разностях был получен статистически значимый (на 5% уровне значимости) отрицательный коэффициент асимметрии (нулевая гипотеза теста – это не отрицательность коэффициента асимметрии), т.е. в среднем наклон на периодах роста более пологий, чем на периодах снижения.

нелинейна по степени влияния. Аналогично для резервных запасов нефти стран ОПЕК, но с обратным влиянием. Но, к примеру, из анализа влияния обменного курса рубля на индекс производства химической отрасли следует: как высокие, так и низкие значения курса оказывали отрицательное влияние на формирование прогнозируемого значения. Возможно, это говорит о том, что существует некоторая предельно убывающая полезность от укрепления рубля для производителей отрасли. При крайне слабом курсе производители не могут закупать импортные факторы производства и средства производства, что заставляет их снижать предложение, а при крайне сильном курсе снижается спрос на продукцию отрасли, как со стороны внешних рынков, поскольку продукция становится относительно дороже, так и со стороны внутреннего рынка, поскольку потребители переключаются на потребление зарубежной продукции, поскольку она становится доступнее.

Для сравнения прогностических качеств модели стека с наилучшей индивидуальной моделью, а также с бенчмарком (безусловная ARIMA-модель), для каждого изучаемого индекса производства используется тест Диболда-Мариано⁹. Для индекса добычи топливно-энергетических ископаемых гипотеза об отсутствии различий между моделями отвергается, положительная статистика теста показывает, что ошибка индивидуальной модели больше ошибки модели стека (обе модели предпочтительнее бенчмарка). Для индекса производства кокса, нефтепродуктов и ядерных материалов гипотеза об отсутствии различий между сетевыми моделями не отвергается, следовательно, модели являются равнозначными (модель стека предпочтительнее бенчмарка). Для индекса химической отрасли гипотеза о равенстве квадратов ошибок отвергается, отрицательная статистика теста показывает, что ошибка модели стека больше ошибки индивидуальной модели, следовательно, индивидуальная нейронная сеть предпочтительнее стека, но при этом точность ARIMA оказывается выше для данной отрасли.

В заключении стоит заметить, что предложенный нейросетевой метод имеет широкий потенциал для улучшения. В первую очередь это касается непосредственно архитектуры и выбора оптимальных гиперпараметров. Однако также чрезвычайно важным является наличие качественной статистики. При доступности более длинных временных рядов и более широкого перечня отраслевых факторов, вполне возможно значительное улучшение качества прогноза и робастности получаемых метрик.

⁹ Нулевая гипотеза теста: различия между прогнозными свойствами двух моделей отсутствуют, т.е. разность среднеквадратических ошибок для двух моделей статистически неотличима от нуля.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Neftici S. N. Optimal prediction of cyclical downturns //Journal of Economic Dynamics and Control. – 1982. – Т. 4. – P. 225-241.
2. Hamilton J.D. (1989), A new approach to the economic analysis of non-stationary time series and the business cycle, *Econometrica*, 57, 357-384.
3. Blatt J. M. On the Frisch model of business cycles //Oxford Economic Papers. – 1980. – Т. 32. – №. 3. – P. 467-479.
4. Burns A. F., Mitchell W. C. Measuring Business Cycles National Bureau of Economic Research //New York. – 1946.
5. Parot A., Michell K., Kristjanpoller W. D. Using Artificial Neural Networks to forecast Exchange Rate, including VAR-VECM residual analysis and prediction linear combination //Intelligent Systems in Accounting, Finance and Management. – 2019. – Т. 26. – №.
6. Chai S. H., Lim J. S. Forecasting business cycle with chaotic time series based on neural network with weighted fuzzy membership functions //Chaos, Solitons & Fractals. – 2016. – Т. 90. – P. 118-126.
7. Al-Maqaleh B. M., Al-Mansoub A. A., Al-Badani F. N. Forecasting using artificial neural network and statistics models //International Journal of Education and Management Engineering. – 2016. – Т. 6. – №. 3. – P. 20-32.
8. Stasinakis C. et al. Forecasting US unemployment with radial basis neural networks, Kalman filters and support vector regressions //Computational Economics. – 2016. – Т. 47. – №. 4. – P. 569-587.
9. Oord A. et al. Wavenet: A generative model for raw audio //arXiv preprint arXiv:1609.03499. – 2016.
10. Borovykh A., Bohte S., Oosterlee C. W. Conditional time series forecasting with convolutional neural networks //arXiv preprint arXiv:1703.04691. – 2017.
11. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. – "Издательский дом Питер", 2017.
12. Chakraborty C., Joseph A. Machine learning at central banks/ Bank of England Working Papers/2017.
13. Tkacz G. et al. Forecasting GDP growth using artificial neural networks/ Bank of Canada/1999.
14. Cook T. R., Macroeconomic Indicator Forecasting with Deep Neural Networks /Federal Reserve Bank Working Paper/2017. .
15. Mitchell T. M. et al. Machine learning. – 1997.

16. Rolnick D., Tegmark M. The power of deeper networks for expressing natural functions //arXiv preprint arXiv:1705.05502. – 2017.
17. Гудфеллоу Я., Иошуа Б., Курвилль А. Глубокое обучение. – Litres, 2018.
18. Makridakis S. et al. The accuracy of extrapolation (time series) methods: Results of a forecasting competition //Journal of forecasting. – 1982. – Т. 1. – №. 2. – P. 111-153.
19. Clemen R. T. Combining forecasts: A review and annotated bibliography //International journal of forecasting. – 1989. – Т. 5. – №. 4. – P. 559-583.
20. Zhang G. P. Time series forecasting using a hybrid ARIMA and neural network model //Neurocomputing. – 2003. – Т. 50. – P. 159-175.
21. Deng L., Yu D. Deep learning: methods and applications //Foundations and trends in signal processing. – 2014. – Т. 7. – №. 3–4. – P. 197-387.
22. Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult //IEEE transactions on neural networks. – 1994. – Т. 5. – №. 2. – P. 157-166.
23. Hochreiter S., Schmidhuber J. Long short-term memory //Neural computation. – 1997. – Т. 9. – №. 8. – P. 1735-1780.
24. C. Olah, «Understanding LSTM Networks» [Интернет-ресурс]. URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>.
25. Fernandez V. Wavelet-and SVM-based forecasts: An analysis of the US metal and materials manufacturing industry //Resources Policy. – 2007. – Т. 32. – №. 1-2. – P. 80-89.
26. Galeshchuk S., Mukherjee S. Deep networks for predicting direction of change in foreign exchange rates //Intelligent Systems in Accounting, Finance and Management. – 2017. – Т. 24. – №. 4. – P. 100-110.
27. Bogaerts T. et al. A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data //Transportation Research Part C: Emerging Technologies. – 2020. – Т. 112. – P. 62-77.
28. Mittelman R. Time-series modeling with undecimated fully convolutional neural networks //arXiv preprint arXiv:1508.00317. – 2015.
29. LeCun Y. et al. Convolutional networks for images, speech, and time series //The handbook of brain theory and neural networks. – 1995. – Т. 3361. – №. 10. – P. 1995.
30. Носко В. П. Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов. – 2004.
31. Hamilton J. D. Calling recessions in real time //International Journal of Forecasting. – 2011. – Т. 27. – №. 4. – P. 1006-1026.

32. Каукин А., Миллер Е. Новая методика Росстата и динамика отраслевых индексов производства в начале 2017 г //Экономическое развитие России. – 2017. – Т. 24. – №. 6.
33. «Росстат» [Интернет-ресурс]. URL: <https://rosstat.gov.ru/>.
34. «Банк России» [Интернет-ресурс]. URL: https://cbr.ru/currency_base/dynamics/.
35. «IMF» [Интернет-ресурс]. URL: <https://www.imf.org/en/Research/commodity-prices>.
36. «EIA» [Интернет-ресурс]. URL: <https://www.eia.gov/finance/markets/crudeoil/supply-opec.php>.
37. Tsay R. S. Nonlinearity tests for time series //Biometrika. – 1986. – Т. 73. – №. 2. – P. 461-466.
38. Lundberg S. M., Lee S. I. A unified approach to interpreting model predictions //Advances in neural information processing systems. – 2017. – P. 4765-4774.
39. Shrikumar A., Greenside P., Kundaje A. Learning important features through propagating activation differences //arXiv preprint arXiv:1704.02685. – 2017.
40. Zhou J., Troyanskaya O. G. Predicting effects of noncoding variants with deep learning-based sequence model //Nature methods. – 2015. – Т. 12. – №. 10. – P. 931-934.
41. Zintgraf L. M. et al. Visualizing deep neural network decisions: Prediction difference analysis //arXiv preprint arXiv:1702.04595. – 2017.
42. «Scimago Journal & Country Rank» [Интернет-ресурс]. URL: <https://www.scimagojr.com/>. [Дата обращения: апрель 2020].
43. Reis J., Santo P. E., Melão N. Artificial intelligence in government services: A systematic literature review //World conference on information systems and technologies. – Springer, Cham, 2019. – P. 241-252.
44. Sharma G. D., Yadav A., Chopra R. Artificial Intelligence and Effective Governance: A Review, Critique and Research Agenda //Sustainable Futures. – 2020. – P. 100004., [Интернет-ресурс].
45. McCarthy J. What is artificial intelligence?. – 1998., [Интернет-ресурс].
46. OECD. COFOG: Classification of the functions of government. Government at a Glance, 194–195., - 2011.
47. «Technologyreview» [Интернет-ресурс]. URL: <https://www.technologyreview.com/2020/03/06/905459/coronavirus-south-korea-smartphone-app-quarantine/>. [Дата обращения: апрель 2020].
48. «Korea Centers for Disease Control and Prevention» [Интернет-ресурс]. URL: <http://www.cdc.go.kr/contents.es?mid=a30301180000>. [Дата обращения: апрель 2020].

49. «Новостное агентство РБК» [Интернет-ресурс]. URL: <https://www.rbc.ru/rbcfreenews/5e8434029a794733f87f1dfd>. [Дата обращения: апрель 2020].
50. «ТАСС» [Интернет-ресурс]. URL: <https://tass.ru/obschestvo/8051195>. [Дата обращения: апрель 2020].
51. «CNBC» [Интернет-ресурс]. URL: <https://www.cnn.com/2020/03/27/coronavirus-surveillance-used-by-governments-to-fight-pandemic-privacy-concerns.html>. [Дата обращения: апрель 2020].
52. «DragonFly» [Интернет-ресурс]. URL: <https://www.globenewswire.com/news-release/2020/04/21/2019221/0/en/Dragonfly-s-Pandemic-Drone-technology-Conducts-Initial-Flights-Near-New-York-City-to-Detect-COVID-19-Symptoms-and-Identify-Social-Distancing.html>. [Дата обращения: апрель 2020].
53. «covid19 hpc consortium» [Интернет-ресурс]. URL: https://covid19-hpc-consortium.org/?mod=article_inline. [Дата обращения: апрель 2020].
54. «ORNL» [Интернет-ресурс]. URL: <https://www.ornl.gov/news/ornl-fight-against-covid-19>. [Дата обращения: апрель 2020].
55. «OLCF ORNL» [Интернет-ресурс]. URL: <https://www.olcf.ornl.gov/>.
56. «C3.ai Digital Transformation Institute» [Интернет-ресурс]. URL: <https://c3di.ai/>. [Дата обращения: апрель 2020].
57. «PSC» [Интернет-ресурс]. URL: <https://www.psc.edu/news-publications/3175-psc-covid-19-update-april-13-2020>. [Дата обращения: апрель 2020].
58. «CSC» [Интернет-ресурс]. URL: <https://www.csc.fi/en/-/koronan-leviamista-supertietokoneella-mallintaneet-tutkijat-tarkeinta-on-nyt-valttaa-vilkkaita-sisatiloja>. [Дата обращения: апрель 2020].
59. «Exscalate4cov» [Интернет-ресурс]. URL: <https://www.exscalate4cov.eu/>. [Дата обращения: апрель 2020].
60. «МГУ им. Ломоносова» [Интернет-ресурс]. URL: https://www.msu.ru/science/main_themes/superkompyuter-mgu-lomonosov-pomozhet-nayti-lekarstvo-priyamoego-deystviya-ot-koronavirusa-.html. [Дата обращения: апрель 2020].
61. «МЦС РАН» [Интернет-ресурс]. URL: <http://www.rscgroup.ru/ru/news/381-superkompyuter-msc-ran-pomozhet-rossiyskim-uchenym-v-sozdanii-medicinskih-preparatov-dlya>. [Дата обращения: апрель 2020].
62. Капелюшников Р. И. Технологический прогресс—пожиратель рабочих мест //Вопросы экономики. – 2017. – №. 11. – Р. 111-140.

63. «МГУ им. Ломоносова» [Интернет-ресурс]. URL: https://www.msu.ru/science/main_themes/superkompyuter-mgu-lomonosov-pomozhet-nayti-lekarstvo-pryamogo-deystviya-ot-koronavirusa-.html.
64. Samuelson P. A. Interactions between the multiplier analysis and the principle of acceleration //The Review of Economics and Statistics. – 1939. – Т. 21. – №. 2. – P. 75-78.
65. Geert Bruinooog, Handbook on Cyclical Composite Indicators For business cycle analysis, Eurostat, 2017.
66. Sichel D. E. Business cycle asymmetry: a deeper look //Economic inquiry. – 1993. – Т. 31. – №. 2. – P. 224-236.
67. De Long J. B. et al. How does macroeconomic policy affect output? //Brookings Papers on Economic Activity. – 1988. – Т. 1988. – №. 2. – P. 433-494.
68. Ball, L., and G. Mankiw. "Asymmetric Price Adjustment and Economic Fluctuations." Manuscript, December 1991.
69. Chetty V. K., Heckman J. J. A dynamic model of aggregate output supply, factor demand and entry and exit for a competitive industry with heterogeneous plants //Journal of Econometrics. – 1986. – Т. 33. – №. 1-2. – P. 237-262.
70. Baldwin R., Krugman P. Persistent trade effects of large exchange rate shocks //The Quarterly Journal of Economics. – 1989. – Т. 104. – №. 4. – P. 635-654.
71. S. N. Neftci Are economic time series asymmetric over the business cycle? //Journal of Political Economy. – 1984. – Т. 92. – №. 2. – P. 307-328.
72. Diebold F. X., Lee J. H., Weinbach G. C. Regime switching with time-varying transition probabilities //Business Cycles: Durations, Dynamics, and Forecasting. – 1994. – Т. 1. – P. 144-165.
73. Filardo A. J. Business-cycle phases and their transitional dynamics //Journal of Business & Economic Statistics. – 1994. – Т. 12. – №. 3. – P. 299-308.
74. Krolzig H. M., Marcellino M., Mizon G. E. A Markov-switching vector equilibrium correction model of the UK labour market //Advances in Markov-Switching Models. – Physica, Heidelberg, 2002. – P. 91-112.
75. Wang J M, Gao T M, McNown R. Measuring Chinese business cycle with dynamic factor models. Journal of Asian Economics, 2009, 20(2): 89–97.
76. Смирнов С. В., Кондрашов Н. В., Петроневич А. В. Поворотные точки российского экономического цикла, 1981–2015 гг //Экономический журнал Высшей школы экономики. – 2015. – Т. 19. – №. 4.
77. Moneta F. (2005), Does the yield spread predict recessions in the Euro area?, International Finance, 8, 2, 263-301.

78. Kauppi H., Saikkonen P. Predicting US recessions with dynamic binary response models //The Review of Economics and Statistics. – 2008. – T. 90. – №. 4. – P. 777-791.
79. Ng E. C. Y. Forecasting US recessions with various risk factors and dynamic probit models //Journal of Macroeconomics. – 2012. – T. 34. – №. 1. – P. 112-125.
80. Dueker M. J. Strengthening the Case for the Yield Curve as a Predictor of US Recessions //Federal Reserve Bank of St. Louis Review. – 1997. – T. 79. – №. 2. – P. 41.
81. Cont R. Empirical properties of asset returns: stylized facts and statistical issues. – 2001.
82. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators //Neural networks. – 1989. – T. 2. – №. 5. – P. 359-366.
83. Fischer T., Krauss C. Deep learning with long short-term memory networks for financial market predictions //European Journal of Operational Research. – 2018. – T. 270. – №. 2. – P. 654-669.
84. Mitchell T. M. Artificial neural networks //Machine learning. – 1997. – T. 45. – P. 81-127.
85. Khashei M., Bijari M. An artificial neural network (p, d, q) model for timeseries forecasting //Expert Systems with applications. – 2010. – T. 37. – №. 1. – P. 479-489.
86. Borovykh A., Bohte S., Oosterlee C. W. Conditional time series forecasting with convolutional neural networks //arXiv preprint arXiv:1703.04691. – 2017.
87. Zhang G. P. A neural network ensemble method with jittered training data for time series forecasting //Information Sciences. – 2007. – T. 177. – №. 23. – P. 5329-5346.
88. Qiu X. et al. Ensemble deep learning for regression and time series forecasting //2014 IEEE symposium on computational intelligence in ensemble learning (CIEL). – IEEE, 2014. – P. 1-6.
89. Bengio Y. Deep learning of representations for unsupervised and transfer learning //Proceedings of ICML workshop on unsupervised and transfer learning. – 2012. – P. 17-36.
90. Sezer O. B., Gudelek M. U., Ozbayoglu A. M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019 //Applied Soft Computing. – 2020. – T. 90. – P. 106181.
91. Hinton G. E. Training products of experts by minimizing contrastive divergence //Neural computation. – 2002. – T. 14. – №. 8. – P. 1771-1800.
92. Kauppi H., Saikkonen P. Predicting US recessions with dynamic binary response models //The Review of Economics and Statistics. – 2008. – T. 90. – №. 4. – P. 777-791.
93. Cao H. et al. Integrated oversampling for imbalanced time series classification //IEEE Transactions on Knowledge and Data Engineering. – 2013. – T. 25. – №. 12. – P. 2809-2822.

94. Liang G., Zhang C. An efficient and simple under-sampling technique for imbalanced time series classification //Proceedings of the 21st ACM international conference on Information and knowledge management. – 2012. – P. 2339-2342.
95. Xu L. et al. Modeling tabular data using conditional gan //Advances in Neural Information Processing Systems. – 2019. – P. 7333-7343.
96. «Indexmundi» [Интернет-ресурс]. URL: <https://www.indexmundi.com/>.
97. «Финам» [Интернет-ресурс]. URL: <https://www.finam.ru/analysis/macroeconomy/>.
98. «Анализ банков РФ» [Интернет-ресурс]. URL: <https://analizbankov.ru/index.php>.
99. Neftici S. N. Optimal prediction of cyclical downturns //Journal of Economic Dynamics and Control. – 1982. – Т. 4. – P. 225-241., [Интернет-ресурс].