



А. С. Каукин, К. А. Касьянова
В. С. Косарев, Е. М. Миллер

**МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ
ПОКАЗАТЕЛЕЙ ОПТОВОГО РЫНКА
ЭЛЕКТРОЭНЕРГИИ РОССИИ
С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ СОВМЕЩЕНИЯ
ДАННЫХ РАЗНОЙ ЧАСТОТНОСТИ**

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Российская академия народного хозяйства и государственной службы
при Президенте Российской Федерации» (РАНХиГС)

Каукин А.С., Касьянова К.А., Косарев В.С., Миллер Е.М.

**МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ ПОКАЗАТЕЛЕЙ ОПТОВОГО РЫНКА
ЭЛЕКТРОЭНЕРГИИ РОССИИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ СОВМЕЩЕНИЯ
ДАННЫХ РАЗНОЙ ЧАСТОТНОСТИ
(препринт)**

Москва, 2021

Аннотация

Целью исследования является разработка новых методов прогнозирования временных рядов с разночастотными данными среди экзогенных факторов; построение прогноза показателей оптового рынка электроэнергии России с использованием методов совмещения разночастотных данных, в том числе на основе алгоритмов сверточных нейронных сетей.

Структура работы представлена четырьмя разделами. В первом разделе анализируются методы прогнозирования временных рядов с совмещением разночастотных данных. Во втором разделе представлена архитектура сверточной сети, позволяющей использовать разночастотные данные. В третьем разделе представлена модель цен на оптовом рынке электроэнергии с использованием данных Атласа российской энергетики. В четвертом разделе представлены рекомендации и основные выводы работы.

Ключевые слова: спрос на электроэнергию, оптовый рынок электроэнергии, генерирующие мощности, рынок на сутки вперед, моделирование цен, разночастотные данные, сверточные нейронные сети.

Abstract

The aim of this study is to develop new methods for forecasting time series with data of different frequencies among exogenous factors; forecasting the indicators of the wholesale electricity market in Russia using methods of combining data of different frequencies, including those based on algorithms of convolutional neural networks.

The structure of the work is presented in four sections. The first section analyzes methods for forecasting time series with combining data of different frequencies. The second section presents the architecture of a convolutional network that allows the use of data of different frequencies. The third section presents a price model for the wholesale electricity market using data from the Atlas of Russian Energy. The fourth section presents recommendations and main conclusions of the work.

Key words: electricity demand, wholesale electricity market, generation capacity, day-ahead market, price modeling, multi-frequency data, convolutional neural networks.

СОДЕРЖАНИЕ

1. Анализ методов прогнозирования временных рядов с совмещением разночастотных данных	4
2. Разработка архитектуры сверточной сети, позволяющий использовать разночастотные данные	7
2.1. Описание архитектуры сверточной сети, позволяющий использовать разночастотные данные.....	16
2.2. Описание теоретических и технических характеристик, обуславливающих возможность работы предложенной сети.....	21
3. Моделирование цен на оптовом рынке электроэнергии (рынка на сутки вперед) и цен балансирующего рынка, в том числе с использованием данных Атласа российской энергетики	26
3.1. Построение прогнозов цен на основе разработанного метода, сравнение результатов с другими методами прогнозирования временных рядов и совмещения размерности	26
3.2. Прогнозирование почасовых объемов спроса на электроэнергию на рынке на сутки вперед с использованием разработанной модели	43
4. Интерпретация полученных результатов.....	61
Список использованных источников.....	65

1. Анализ методов прогнозирования временных рядов с совмещением разночастотных данных

Большинство работ по моделированию и прогнозированию временных рядов предполагает, что процессы, порождающие данные, характеризуются одинаковой частотой. При анализе данные с более высокой частотой агрегируются до самой низкой частоты путем усреднения, суммирования или взятия репрезентативного значения (например, первый месяц года), либо, наоборот, данные с более низкой частотой интерполируются до самой высокой частоты. Однако, при агрегации теряется информация о структуре временного ряда, что приводит к ухудшению качества модели. При использовании обычных методов интерполяции, как правило, не учитывается вся доступная информация о выборке.

Помимо проблемы использования разночастотной выборки, существует проблема, заключающаяся в том, что многие экономические временные ряды публикуются с задержкой или могут пересматриваться. Многие макроэкономические временные ряды предоставляют своевременную информацию о текущем состоянии экономики, однако публикуются с задержкой и достаточно большой периодичностью. Таким образом, существует потребность в моделях, учитывающих частотность данных.

Изучение методов работы с наборами данных, содержащих ряды с разной периодичностью (ежеквартальной, ежемесячной, ежедневной), необходимо, поскольку интерес к прогнозированию таких рядов существует во многих областях науки. Использование последних доступных данных, часто поступающих с разной частотой, имеет решающее значение для своевременного и статистически эффективного обновления прогнозов. В частотности, при решении задач по планированию государственной экономической политики, экономико-финансовом анализе, решении задач, связанных с эпидемиологией и прогнозированием погоды и т.д. был разработан широкий спектр доступных методов, причем по мере возникновения новых потребностей и появления новых типов данных они развиваются и появляются новые подходы.

Одними из первых попыток разработки моделей, учитывающих данные с разной частотностью, были модели с уравнением связи (linkage function models, LF-модель) и уравнения «моста» (bridge function models, BF-модель). В LF-моделях генерируются прогнозы разной частоты, затем прогнозы объединяются для повышения точности прогнозов низкочастотных временных рядов. BF-модели же фактически являются моделями с рядами одной частоты. Высокочастотные ряды прогнозируются до желаемого горизонта прогноза по от-

дельной модели временных рядов. Далее, эти прогнозы агрегируются до более низкой частоты и включаются в модель временных рядов с более низкой частотой (как дополнительные регрессоры). Уравнения моста особенно полезны для наукастинга – к предсказанию настоящего, ближайшего будущего и совсем недавнего прошлого на основе доступной информации, наблюдаемой с более высокой частотой.

Популярным подходом к моделированию временных рядов разной частотности является подход, предложенный Задрозным [1], для оценки модели векторной авторегрессии со скользящим средним (VARMA) с временными рядами различной частоты. При данном подходе предполагается, что данные порождаются с самой высокой частотой, однако не обязательно наблюдаются с самой высокой частотой, но могут использоваться для получения прогноза любой переменной с этой частотой. Переменные, которые наблюдаются с более низкой частотой, считаются периодически пропущенными.

В статье [2] было показано, что качество прогноза ВВП США может быть улучшена при использовании векторной модели пространства-состояний VARMA по сравнению с авторегрессионными бенчмарк-моделями.

Другим подходом к анализу временных рядов разной частотности, является так называемый MIDAS подход или подход смешанной выборки (MIXed DATA Sampling), предложенный в статье [3]. В модели MIDAS условные математические ожидания ряда моделируются как распределенные лаги регрессоров на более высоких частотах. Фактически ряды низкой частоты регрессируются на более высокочастотные ряды. Чтобы избежать большого числа параметров, используется функция взвешивания. Подход MIDAS позволяет учитывать наблюдаемые шоки и использует полиномы распределенных лагов запаздывания для получения спецификаций модели с небольшим числом регрессоров.

Чтобы преодолеть эти недостатки методов агрегирования и интерполяции, помимо моделей в пространстве состояний, моделей смешанной выборки MIDAS, используется также байесовский подход [4]. В этой статье для оценки MF VAR (mixed frequency VAR) модели для заполнения пропущенных наблюдений используется байесовский подход вместо модели пространства состояний.

Преимущество подхода смешанной выборки по сравнению с альтернативными, такими как модели пространства состояний и MF-VAR, в которых используется фильтр Калмана, состоит в том, что MIDAS модели более лаконичные и менее чувствительны к ошибкам спецификации за счет использования нелинейных полиномов запаздывания [5]. Ниже в таблице 1 представлены краткие характеристики различных подходов к работе с данными разной частотности (модели, использующие фильтр Калмана, модели смешанной выборки

MIDAS и модели, основанные на байесовском подходе), их основные преимущества и недостатки.

Таблица 1

Сравнение основных подходов, позволяющих работать с наборами данных с рядами со смешанной частотностью

Фильтр Калмана	Модели смешанной выборки MIDAS	Байесовский подход
Модели с динамическими ненаблюдаемыми факторами (Dynamic Latent Factor Models) и модели VAR со смешанной частотностью	Нелинейная модель наименьших квадратов с распределенными лагами	Позволяет использовать априорную информацию, включать в модель прошлую информацию о параметрах.
Модели имеют представление в форме пространства состояний, к которой применяется фильтрация	Вычислительно более податливая (tractable), т. е. для получения оценок необходимо относительно немного времени.	Когда становятся доступными новые наблюдения, предыдущее апостериорное распределение можно использовать как априорное.
Лаконичные модели (небольшое число параметров)	Необходим значительно больший объем данных	Позволяет получать достаточно точные выводы, даже при небольшой выборке.
Зависят от гораздо меньшего набора данных, который необходимо регулярно обновлять		Высокие вычислительные затраты, особенно для моделей с большим количеством параметров.
Используются технически и вычислительно сложные модели, особенно для данных со смешанной частотой		

Примечание – Источник: составлено авторами.

2. Разработка архитектуры сверточной сети, позволяющий использовать разночастотные данные

Научная литература по прогнозированию цен и спроса на электроэнергию в части применяемых методов весьма разнообразна. Верон в своем исследовании [6] выделяет пять направлений: многоагентные, фундаментальные, методы приведенной формы, статистические и методы машинного обучения. Мультиагентные модели имитируют работу энергетического рынка и моделируют одновременно стороны спроса и предложения. Фундаментальные или структурные методы ориентированы на изучение влияния физических и экономических факторов на цены на электроэнергию и произведенные объемы. Как правило, в этом случае применяются достаточно хорошо интерпретируемые эконометрические линейные модели. К методам приведенной формы относятся марковские модели с переключениями режимов и модели диффузионно-скачкообразных процессов. Данный тип моделей имеет преимущества при наличии тяжелых хвостов у распределения целевой переменной. К статистическим методам, как и к фундаментальным, могут относиться как линейные эконометрические модели, так и более сложные нелинейные. Данное направление от фундаментального отличается целью использования математического аппарата. В последнем случае целью является построение качественного прогноза. Модели машинного обучения стоят в этом же ряду. В сфере экономики электроэнергетики они приобрели особую популярность [7,8,9,10,11,12,13]. Во многом это связано со специфическими особенностями данных, среди которых в первую очередь стоит отметить большую размерность и нелинейность отношений. Для этих условий модели машинного обучения подходят в наибольшей степени [14]. С развитием направления искусственных нейронных сетей все больше исследователей экспериментируют с различными архитектурами, приспособивая их под специфические особенности энергетического рынка с целью превзойти бенчмарк при прогнозировании цен или спроса на электроэнергию. К примеру, в работе [8] используется сверточная нейронная сеть, основные особенности архитектуры которой были впервые описаны в работе [15]. Авторы данной работы разработали свою архитектуру для генерации аудиосигнала. При этом сама идея сверточных нейронных сетей восходит к работам Яна Лякуна [16], который создал данный тип нейросетей, тем самым совершив революцию в компьютерном зрении. В настоящее время сверточные сети продолжают активно развиваться и достигать новых возможностей для человека в области анализа фото и видео изображений. Рекуррент-

ные нейронные сети также широко используются для прогнозирования показателей энергетического рынка и при этом находятся уже в более естественной для себя области [13]. В данном направлении стоит отдельно отметить работы отечественных ученых [17,18].

Рекуррентная сеть – это специальный тип нейронной сети, используемый для генерации временных рядов или последовательных данных (текст, звук и т.д.). Их особенностью является возможность связывать предыдущую информацию с текущей. Для прогнозирования цен на электроэнергию также применялись генеративные нейронные сети [19], нейронные сети с нечеткой логикой [12]. Отдельного внимания заслуживают гибридные модели. В работах [10,9] используются комбинации сверточных и рекуррентных нейронных сетей. Такой подход позволяет использовать преимущества обоих видов архитектур при работе с временными последовательностями.

Стоит обратить внимание на доминирующее среди авторов вышеприведенных работ мнение относительно того, что прогнозирование цен и спроса на электроэнергию представляет собой многомерную задачу прогнозирования временных рядов [13,11,10]. В работах [20,21], демонстрируется, что модели с дополнительными объясняющими переменными дают лучшие результаты прогнозирования, чем одномерный подходы.

Однако показатели энергетического рынка, в том числе цены и объемы произведенной электроэнергии, могут иметь высокую частоту измерения – часы и минуты. И, соответственно, прогноз, имеющий практическую ценность для участников рынка, должен быть построен с таким же уровнем детализации. При этом большая часть причинных экономических факторов имеет меньшую частоту измерения – месячную (индексы промышленного производства) или квартальную (валовый региональный продукт). Таким образом, возникает проблема совмещения данных разной частотности. В указанных выше работах эта проблема оставалась в стороне, поскольку авторы в своих моделях использовали в качестве объясняющих переменных данные одной размерности или заранее приведенные к таковой. К примеру, в работах [7,22,23] для прогнозирования цен/спроса на электроэнергию в качестве эндогенных факторов использовались погодные условия и синтетические переменные, сконструированные для учета выходных и праздничных дней, а также учета сезонности. В работе [10] для прогнозирования электропотребления среди объясняющих переменных использовались данные о активной и реактивной мощности электроэнергии и силе тока, имеющие одну размерность с объясняемой переменной. Однако в работе [24] указывается, что национальные показатели экономической деятельности и финансового рынка имеют значимое влияние на цены и объемы производства электроэнергии. И к тому же являются причинами для этих показателей.

Таким образом, к недостаткам работ [9,10,11,13] следует отнести возможную неполноту моделей, поскольку среди объясняющих факторов отсутствовали показатели экономической активности населения. Однако представляется, что нейросетевые методы, лежащие в основе данных исследований, могут быть модифицированы для работы с разночастотными данными. В частности, сверточная сеть позволяет данные не только сворачивать, но и из меньшей размерности переводить в большую. Эта особенность может позволить использовать в числе объясняющих переменных разноразмерные данные без необходимости приводить их к единой структуре путем простых математических операций, таких как усреднение или интерполяция, которые могут лишить данные части значимой информации или же напротив, привнести искажающий шум.

В собственной модели мы предполагаем использовать связку сверточных (CNN) и рекуррентных нейронных сетей (RNN). При этом блок сверточной сети будет разработан с учетом возможности в качестве одного паттерна, т.е. обучающего образца принимать разночастотные данные. Следуя опыту предшественников, мы предполагаем, что сверточный блок в нейронной сети будет способствовать удалению шума и учету пространственной корреляции между факторами, а также, как было сказано, позволит комбинировать разноразмерные факторы. Рекуррентный блок предназначен для извлечения полезной для прогнозирования информации из временной структуры данных.

Комбинация CNN-RNN использовалась для прогнозирования цен и спроса на электроэнергию в работах [9] [10]. К тому же, как отмечается в работе [10], в последние годы подобная архитектура, оправдала себя во многих областях, где для прогнозирования требовалось извлечение одновременно временных и пространственных характеристик многомерных временных рядов. К примеру, в работе [25] было продемонстрировано, что модель голосового поиска, разработанная на базе данной архитектуры, имеет высокую устойчивость к шуму. В работе [26] модель, объединяющая сверточную сеть и двунаправленную рекуррентную сеть, показала высокую эффективность при распознавании действий человека на видео. В области медицины подобный подход позволил разработать модель для точного обнаружения аритмий на ЭКГ [27]. Сверточная двунаправленная рекуррентная сеть в работе [28] была использована для решения задачи прогнозирования износа инструмента.

Возвращаясь к задаче прогнозирования показателей электроэнергетического рынка, стоит более подробно проанализировать опыт построения и обучения модели в работе [10], поскольку результат экспериментов показал, что предлагаемая авторами модель обеспечивает более низкие ошибки прогноза (MSE, RMSE, MAE и MAPE) в сравнении с рядом традиционных архитектур, таких как: LSTM, GRU, Bi-LSTM и Attention LSTM.

Все бенчмарки являются нетривиальными нейросетевыми моделями, специально разработанными для моделирования и прогнозирования временных рядов. LSTM – является усовершенствованной версией простой рекуррентной сети, в которой за счет наличия вектора долгосрочной памяти решена проблема исчезающих градиентов. GRU, в свою очередь, вариация LSTM, которая отличается от прародителя меньшим числом обучаемых параметров, но при этом также эффективна. Bi-LSTM – это двунаправленная рекуррентная сеть, хорошо зарекомендовавшая себя в задачах машинной лингвистики и, в частности, машинном переводе. Attention LSTM – это рекуррентная нейронная сеть совмещенная с сетью внимания. Сеть внимания позволяет придавать различные веса выходам ячейки LSTM, формируя таким образом взвешенный прогноз из множества состояний рекуррентного нейрона.

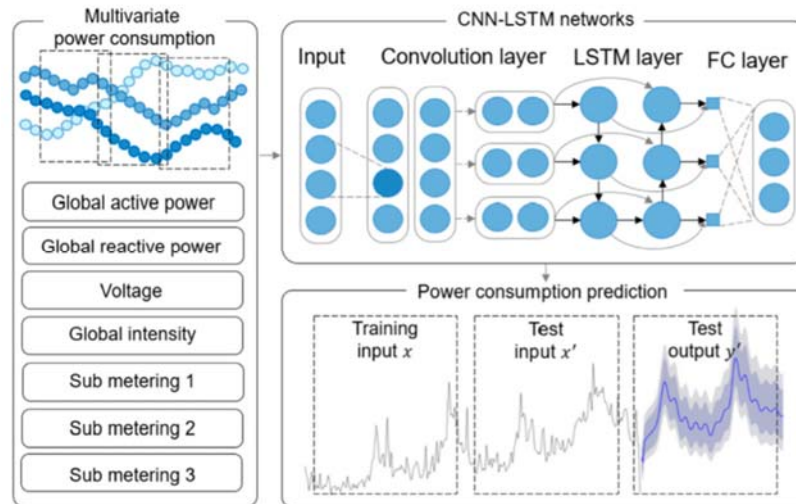
Таким образом, полученный результат в работе [10] заслуживает внимания. В этой статье авторы используют нейронную сеть CNN-LSTM для прогнозирования потребления электроэнергии в жилых домах. Анализируемые данные состоят из 2 млн. наблюдений и включают 12 переменных. В наборе данных собраны поминутные показатели потребления электроэнергии за четыре года (с 16 декабря 2006 г. по 26 ноября 2010 г.) в жилых домах во Франции, оснащенных индивидуальными счетчиками учета.

На рис. 1 показана общая архитектура предложенной авторами сети. Модель состоит из последовательного соединенных сетей CNN и LSTM. Сеть принимает на вход матрицу размерности 60×12 , где 60 – это наблюдений во времени, а 12 – это число факторов, которые влияют на потребление электроэнергии. Среди них: активная и реактивная мощность, напряжение, сила тока, дата и время. Прогнозируемая переменная – это активная мощность электроэнергии, измеряемая в киловатт-часах.

CNN состоит из двух сверточных слоев с ядром свертки равным двум и единичным шагом. В качестве функции активации после операций свертки используется «ReLU». За слоями свертки следуют «Pooling» слои, имеющие аналогичные характеристики фильтров. Фильтр «Pooling» слоя не имеет обучающихся параметров и используется для снижения размерности, сохраняя максимальное значение из каждого кластера нейронов предыдущего слоя. Затем слои объединяются и далее следует рекуррентная сеть.

Связующим слоем между CNN и LSTM служит специальный сверточный слой – «TimeDistributed». Он преобразует данные из многомерной размерности, получаемые на выходе из предыдущего слоя сверточной сети в одномерный вектор. При этом операция свертки применяется ко всем элементам тензора, относящимся к одному временному периоду с единичными весами, т.е. выполняется простое сложение по временной оси. Сеть LSTM состоит из 64 ячеек. Каждая ячейка использует данные, выдаваемые промежуточным

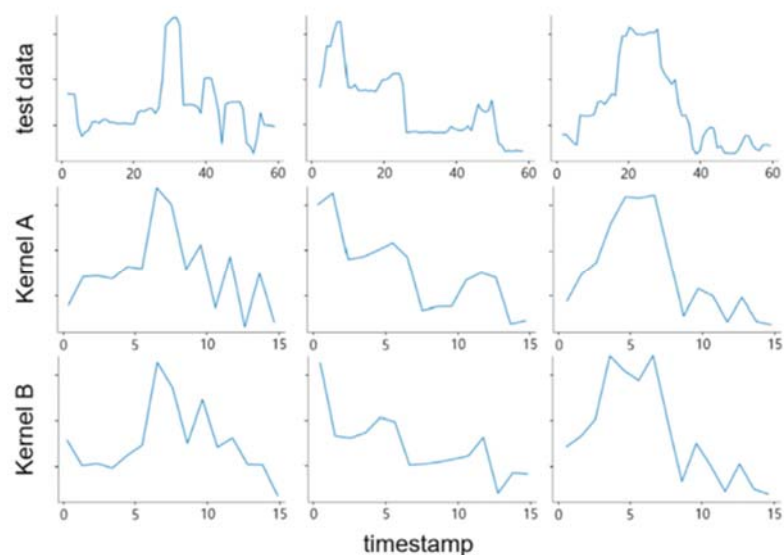
слоем. Особенность LSTM ячейки состоит в наличии долгосрочной памяти. Так называемый вектор памяти может хранить информацию о важных характеристиках потребления электроэнергии, извлеченную через сверточные слои сразу из нескольких паттернов. Эта особенность позволяет учитывать при построении прогноза долгосрочные зависимости в данных.



Примечание: Input – входной слой; Convolution layer – сверточный слой; LSTM layer – рекуррентный слой; FC layer – полносвязный слой; Global active/reactive power – активная и реактивная мощности электроэнергии (кВт); Voltage – вольт; Global intensity – сила тока (ампер); Sub metering 1,2,3 – Данные датчиков потребления (Ватт-час). Источник [10].

Рис. 1. Архитектура гибридной модели сверточной и рекуррентной нейросети для прогнозирования спроса на электроэнергию

Кроме того, в работе Ким и Че [5] представляется полезным с точки зрения понимания принципа работы предложенной сети, проведенный анализ внутренних механизмов. К примеру, на рис. 2 представлены результаты обработки данных активной мощности в результате операций свертки и объединения.

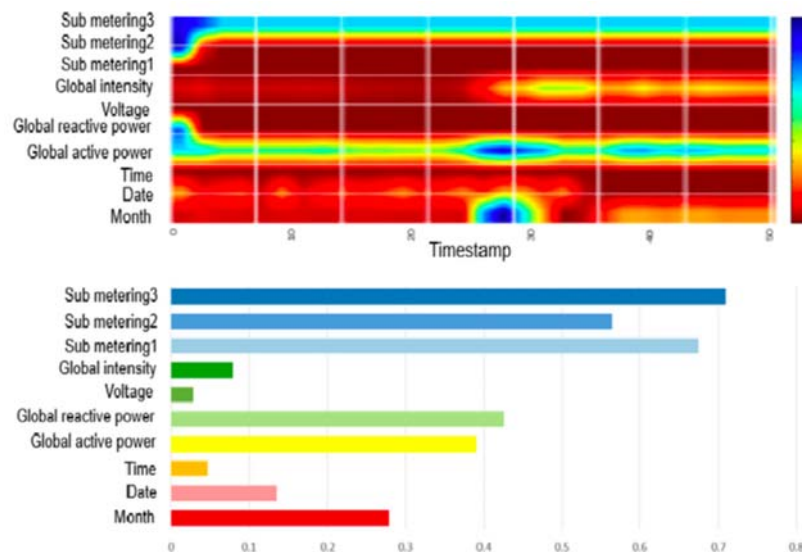


Примечание – Источник [10].

Рис. 2. Визуализация работы сверточных слоев нейронной сети

Можно видеть, как 60-минутное временное окно показателей потребления электроэнергии уменьшается в четыре раза за счет двух слоев свертки и объединения. При этом шум уменьшается, но локальные и глобальные особенности сохраняются. Выходной результат этого слоя используется в качестве входных данных для LSTM. Таким образом, авторы демонстрируют, что предложенная сеть CNN-LSTM выполняет моделирование с минимальной потерей информации.

Также авторы показывают способ ранжирования факторов по влиянию на прогноз. На рис. 3 на верхнем графике показана тепловая карта активаций последнего сверточного слоя для некоторого паттерна. На нижнем графике показана диаграмма, столбцы которой представляют усредненные два раза значения результатов активационных функций: сперва усреднение происходит для всех паттернов для одного и того же фактора, а затем по времени. В результате получается важность признака.



Примечание: Global active/reactive power – активная и реактивная мощности электроэнергии (кВт); Voltage – вольт; Global intensity – сила тока (ампер); Sub metering 1,2,3 – Данные датчиков потребления (Ватт-час). Источник [10].

Рис. 3. Визуализация последнего сверточного слоя нейронной сети

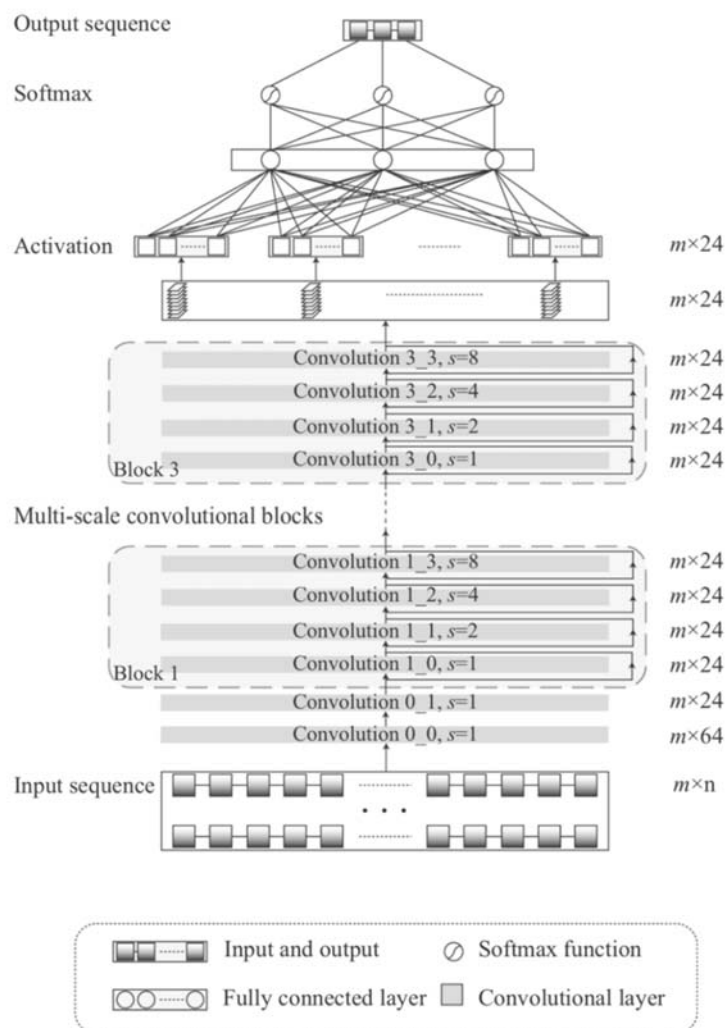
На тепловой карте чем больше синего цвета, тем больше влияние на предсказание, а чем больше красного, тем меньше влияние на предсказание. Таким образом, можно видеть, что факторы sub_metering1, sub_metering2 и sub_metering3 являются важными переменными, оказывающими значительное влияние на формирование прогноза. При этом авторы отмечают, что результат визуализации – это воспроизведение отношений одного рецептивного поля и сверточного фильтра. Однако в сети, как правило, на последнем слое множество фильтров и рецептивных полей, к которым данные фильтры применяются. Внутренние рецептивные поля – это результат работы сверточных фильтров предыдущих слоев. В таком случае важность признаков может быть рассчитана и визуализирована путем сложения результатов. Впрочем, стоит отметить, что для этих целей предусмотрена 1*1 свертка, которая может быть использована, как внутри основной сети, так и специально для данной задачи. Подобная внутренняя визуализация полезна для понимания поведения модели. В практической части собственного исследования мы также будем прибегать к анализу промежуточных выходных данных сконструированной нейронной сети.

Еще одной работой, практические наработки которой также будут использованы при разработке собственной модели, является исследование китайских ученых [8]. Авторы данного исследования предложили использовать сверточную нейронную сеть особенной конструкции с вниманием для прогноза цен на оптовом рынке электроэнергии в Канадском городе Онтарио. В качестве входных данных используются объемы генерации, мощности,

лаговые значения цены энергосистемы Онтарио. Выходные данные – это прогноз пиковой, минимальной и нормальной цены на сутки вперед с 5-ти минутным интервалом. Выборка включает в себя данные 5-минутной частоты измерения за 2021–2015 гг. энергосистемы Онтарио.

К особенностям данной работы стоит отнести непосредственно архитектуру нейронной сети, которая имеет расширенные свертки, позволяющие учитывать временные долгосрочные зависимости в данных, а также надстройку в виде сети внимания. Первая особенность архитектуры сети позволила авторам отказаться от использования рекуррентных ячеек, что существенным образом позволило уменьшить количество обучаемых параметров и снизить вычислительную сложность. Использование механизма внимания позволило увеличить точность прогнозирования. Как указывалось в оригинальной работе [29], механизм внимания позволяет выделять наиболее значимые для прогнозирования шаблоны среди скрытых состояний рекуррентной нейронной сети. Позже идея данного механизма была адаптирована и применена для сверточных нейронных сетей, в том числе, и в рассматриваемой работе. На рисунке 4 приведена архитектура описанной нейронной сети.

Авторы работы указывают, что обычная сверточная операция извлекает признаки только в окрестностях, что отражает ограниченные отношения между нейронами. При использовании расширенных, или так называемых причинных сверток соседние элементы в ядре расширенной свертки могут соответствовать несмежным нейронам на карте признаков. Операция расширенной свертки расширяет рецептивное поле ядра свертки и позволяет извлекать признаки больших масштабов, с использованием различных шагов расширения. При этом важно заметить, что увеличения количества параметров не происходит. Это возможно благодаря архитектуре сверточного блока нейронной сети, изображенной на рис. 4. Идея данной архитектуры впервые появилась в работе [15]. Ее авторы изучали способы генерации звуковых сигналов. Предложенная концепция оправдала себя и стала альтернативой рекуррентным сетям с долгосрочной памятью.

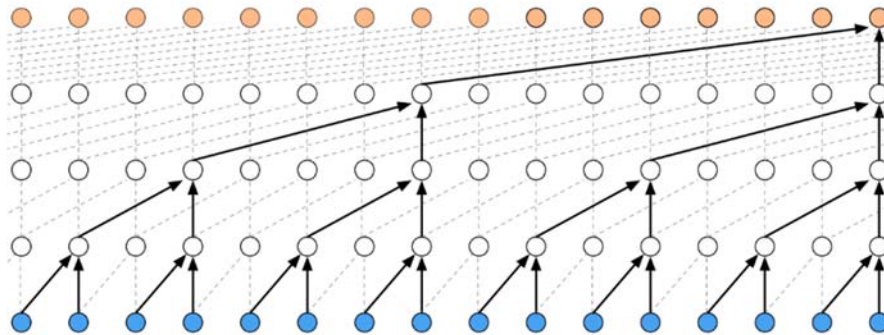


Примечание: Input and output - входные и выходные слои сети; Fully connected layer – полносвязный слой нейронной сети; Softmax function – функция обобщенной сигмоиды, применяемая для многоклассовой классификации; Convolution layer – сверточный слой нейронной сети. Источник [8].

Рис. 4. Архитектура сверточной нейронной сети для прогнозирования спроса на электроэнергию

В Keras реализовать представленную на изображении архитектуру расширенной свертки можно, используя одномерное сверточное ядро (conv1d) со смещением на 2 шага (strides=2) на первом уровне и, сохраняя размерность карты сверточного слоя (padding=«causal»), применяя одномерную свертку (conv1d) с ядром 2 и шагом внутри ядра 2 (dilation_rate=2), а также смещением ядра на четыре шага (strides=4) и так далее. При этом шаг внутри ядра применять необязательно, поскольку ячейки без значений в результате операции padding=«causal» заполняются нулями и, следовательно, свертка с ядром три и

шагом четыре приведет к аналогичному результату. Однако последний вариант предполагает присвоение некоторого веса нулевому элементу при начальной инициализации параметров и, следовательно, участие этой связи в расчете градиента. Такая конструкция увеличивает вычислительную сложность, хотя и приводит к аналогичному результату. Возможен и альтернативный вариант без сохранения размерности карт. В таком случае можно использовать ядра одинаковой размерности, т.е. без пропуска значений внутри на каждом уровне. При этом смещение ядра на каждом уровне, кроме последнего, должно быть равно двум.



Примечание: источник [15].

Рис. 5. Визуализация принципа работы причинной сверточной сети

Описанные преимущества рассмотренных выше архитектур предполагается использовать при разработке собственной модели. В качестве базовой модели выступит нейронная сеть, состоящая из трех блоков: сверточной сети, рекуррентной, а также сети внимания. При этом особенностью сверточной сети станет возможность принимать в качестве входа матрицы разных размерностей, что позволит в числе экзогенных переменных одновременно использовать данные разной частотности, к примеру, месячной и дневной. Отдельные блоки нейронной сети, как и их внутреннее устройство, можно рассматривать, как гиперпараметры. По этой причине, в практической части работы для определения оптимальной структуры конечной сети параметры блоков предполагается изменять в зависимости от тестовых результатов.

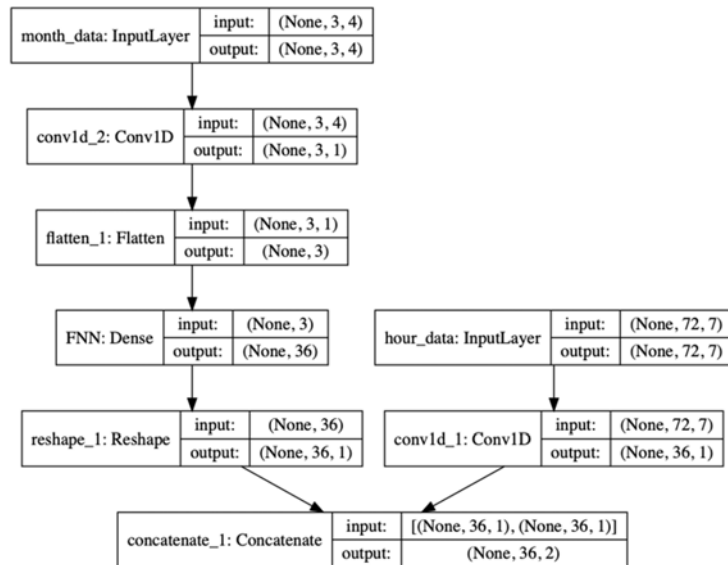
2.1. Описание архитектуры сверточной сети, позволяющий использовать разночастотные данные

Предлагаемая в данной работе архитектура нейронной сети для прогнозирования показателей энергетического рынка представляет собой комбинацию сверточной (CNN) и рекуррентной сети (LSTM) с механизмом внимания. Возможность использования разночастотных данных в такой сети обусловлена специальной конструкцией сверточного блока. Идея состоит в том, что данные с разной частотой измерения подаются в сеть независимо в

виде двух отдельных матриц. К каждой матрице затем применяются операции, сводящие данные к одной размерности при помощи аффинных преобразований с обучаемыми параметрами. Первая матрица имеет размерность $p_1 * m_1$, где p_1 – это гиперпараметр, отвечающий за продолжительность временного окна высокочастотных данных, а m_1 – число факторов. Протяженность данного окна определяет протяженность рецептивного поля, а количество факторов – его ширина для применяемого в последующем сверточного слоя, что и есть аффинные преобразования для матрицы с высокочастотными данными. В результате размерность высокочастотных данных внутри сети сокращается. Масштаб сокращения размерности определяется внутренними параметрами сверточного слоя. Вторая матрица имеет размерность $p_2 * m_2$, где p_2 – это продолжительность временного окна для низкочастотных данных, а m_2 – число факторов. С матрицей низкочастотных данных совершается обратная операция – увеличение размерности от p_2 до p_1 по оси времени. Далее возможны два сценария.

Первый – это использовать полносвязную нейронную сеть в качестве декодировщика, приводящего размерность вектора с низкочастотными данными к редуцированной размерности вектора высокочастотных данных. Применение ANN в данном случае – это аффинные преобразования для месячных данных. Подобная конструкция предназначена для проверки следующей гипотезы: экономические показатели с частотой наблюдения в один месяц могут улучшить долгосрочный прогноз цен/объемов электроэнергии на оптовом рынке. Под долгосрочным прогнозом понимается почасовой прогноз до трех дней. Данная гипотеза основана на результатах, полученных ранее в исследовании: в модели «Prophet» экономические факторы имеют статистически значимое значение. Таким образом, предполагается, что высокочастотные данные в нейросетевой модели служат в первую очередь источником информации относительно поведения случайной компоненты и, в том числе, тренда, а экономические факторы – источник информации только для определения положения тренда. При этом сеть прямого распространения, инкорпорированная внутри первого блока общей нейросети, выступает в роли условной сети внимания, хотя по определению таковой и не является. Полноценная сеть внимания будет находиться в конце предлагаемой архитектуры над рекуррентным блоком. В данном случае имеется в виду, что можно провести некоторую аналогию в функциональности. Задачей полносвязной подсети является отображение вектора низкочастотных данных (к примеру, месячных) в вектор большей размерности, равный редуцированному вектору высокочастотных данных (к примеру, почасовых). Поскольку сеть полносвязная, то каждый элемент нового вектора – это взвешенная сумма всех элементов месячных данных, пропущенная через функцию активации. Веса определяют, какие лаговые значения месячных данных и в какой степени влияют

на формирование целевого вектора, который в дальнейшем поэлементно складывается с вектором-представлением часовых данных, а полученный результат используется последующими блоками непосредственно для построения прогноза. Необходимость сжатия всей информации в один вектор в конце первого блока приводит к поиску оптимальных параметров внутренней полносвязной сети в процессе минимизации ошибки отклонения. Стоит отметить, что вместо полносвязной сети может быть использована и полноценная сеть внимания, однако на данном этапе это представляется излишним и может быть протестировано в дальнейшем. На рис. 6 приведена архитектура первого блока по описанной схеме.



Примечание – Источник: построено авторами с использованием пакета «Keras» для «Python»

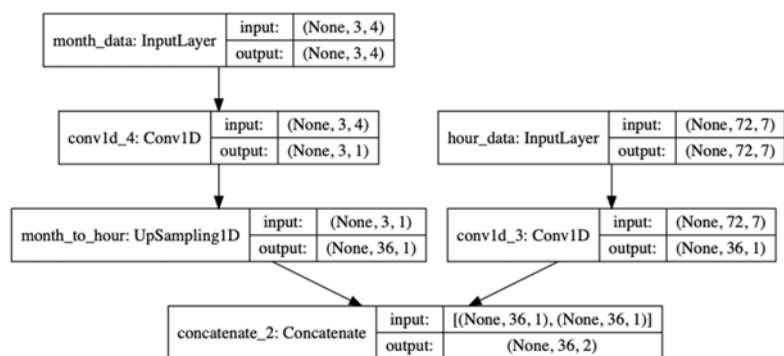
Рис. 6. Конструкция сверточного слоя нейронной сети для обработки разночастотных данных с внутренней полносвязной сетью

Иллюстрация для вышеприведенной схемы взята из предварительно опробованной на практике собственной модели. Поэтому схема содержит подробности относительно размерностей входных данных. К примеру, в процессе опытов было использовано 4 макроэкономических фактора с временным окном, равным трем месяцам и 7 факторов с почасовым разрешением и историей наблюдений в три дня. Отсюда размерность матрицы месячных данных 3×4 , а почасовых 72×7 . Значение «None» ставится по умолчанию и обозначает третью размерность – число образцов, доступных для обучения нейронной сети. Подробнее об этом речь пойдет позже, непосредственно в практической части работы. На данном тапе обратим внимание на то, как трансформируются размерности матриц и векторов с данными внутри приведенного блока. На втором слое для месячных данных используется операция свертки с единичным фильтром, т.е. на выходе из матрицы получается вектор из трех эле-

ментов. Каждый элемент нового вектора – это взвешенная сумма макроэкономических показателей в один момент времени. Веса – это обучаемые параметры нейронной сети. Такая операция на первом слое необходима для того, чтоб переводить в большую размерность не четыре вектора, для чего потребовалось бы построить четыре независимых нейронных сети, а один. Таким образом, вектор на выходе из второго слоя – это главная компонента четырех факторов, который поступает на вход в третий слой, именуемый на схеме ANN – полно-связная нейронная сеть. Промежуточный слой «Flatten» является техническим без каких-либо параметров и необходим для преобразования представления данных. На выходе из третьего слоя по вышеописанной процедуре получается вектор размерности 36. Такой же размерности получается вектор в результате преобразования сверточным фильтром с ядром два и шагом два почасовых данных. В последнем слое данного блока происходит конкатенация полученных векторов по временной оси.

Второй способ отображения низкоразмерных данных в пространство более высокой размерности проще и не требует дополнительных параметров. К вектору месячных данных может быть применена операции «UpSampling1D», которая повторяет каждый элемент первоначального вектора с месячными данными по оси времени некоторое количество раз, приводя данные к заданной размерности. Такая операция аналогична ручному заполнению пропусков последним известным значением. Однако ее применение внутри нейронной сети приводит к необходимости не увеличивать размерность месячных данных для соответствия с часовыми, в результате чего одно месячное наблюдение повторялось бы 744 раза, а увеличивать размерность до уже сжатого вектора часовых наблюдений в результате применения операции свертки. На рисунке 7 представлена архитектура первого блока нейронной сети.

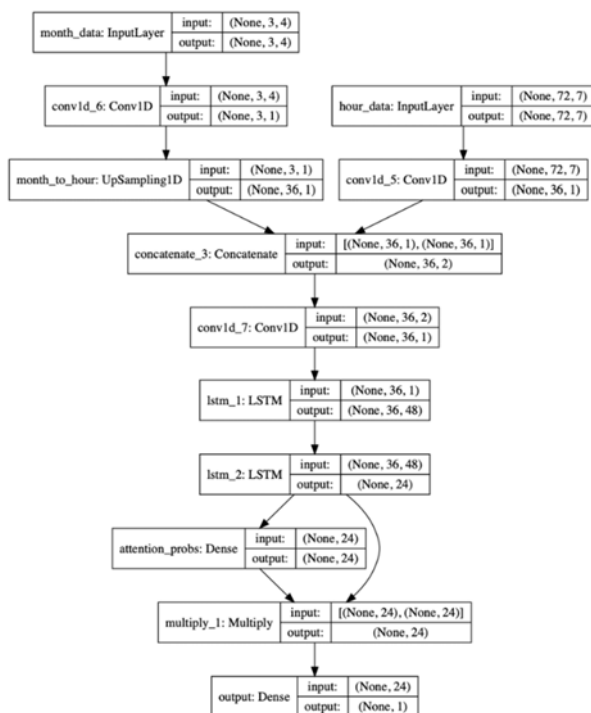
Как было сказано, размерность временного окна как матрицы с высокочастотными, так и с низкочастотными данными – это гиперпараметры. Они могут быть определены, как на основе тестов: сравнения точности прогноза на тестовой выборке аналогичной по устройству нейронной сети, но с разными окнами, так и определены экспертно, поскольку несут в себе определенный экономический смысл. Продолжительность временного окна – это число лаговых значений эндогенных факторов, обуславливающих вывод нейронной сети.



Примечание – Источник: построено авторами с использованием пакета «Keras» для «Python»

Рис. 7. Конструкция сверточного слоя нейронной сети для обработки разночастотных данных с механизмом «UpSampling1D»

Приведенные к одной размерности месячные и часовые данные используются совместно для выделения значимых паттернов при прогнозировании непосредственно рекуррентной нейронной сетью. На рисунке 8 приведена полная архитектура нейронной сети с использованием второго подхода.



Примечание – Источник: построено авторами с использованием пакета «Keras» для «Python»

Рис. 8. Архитектура собственной нейронной сети для прогнозирования почасовых цен на электроэнергию на оптовом рынке

После блока конкатенации следует не рекуррентный, а сверточный слой с одномерной сверткой в пространстве. Заметим, что на выходе временная размерность сохраняется.

Необходимость такой операции объясняется, как и в случае предобработки месячных данных перед полносвязным слоем, снижением размерности. Далее вектор передается в блок рекуррентной нейронной сети – последовательную связку двух ячеек LSTM. В процессе обучения в данном блоке главным образом определяются временные закономерности, а также строится отображение входного вектора в отдельные пространства с уникальными паттернами. После рекуррентного блока полученный вектор передается в сеть внимания. Сеть внимания определяет какой из выделенных паттернов окажет большее влияние на построение прогноза в определенный момент времени. Затем данные попадают в последний выходной слой нейросети, который представляет сумматор с линейной функцией активации. Выход нейронной сети – это цена покупки электроэнергии в определенный час. Для долгосрочного прогноза выход может быть расширен до необходимого количества прогнозных значений.

2.2. Описание теоретических и технических характеристик, обуславливающих возможность работы предложенной сети

Выбранная в качестве основной модели гибридная нейросеть в своей структуре имеет последовательно связанные сверточную (CNN) и рекуррентную (LSTM) нейронные сети. Идея применения CNN к обработке временных рядов тесно связана с дискретным вейвлет преобразованием [30]. Вейвлет преобразование позволяет разложить временной ряд на ортогональные компоненты на разных частотах, которые локализованы во времени. Непосредственно технику вейвлет преобразования также используют для прогнозирования временных рядов [31]. И в том числе в стеке с рекуррентными нейронными сетями, т.е. вместо блока сверточных нейронных сетей [32]. Декомпозиция сигнала на частотные компоненты происходит со сжатием данных при помощи высокочастотных и низкочастотных фильтров h (уравнение (1)) и g (уравнение (2)):

$$h(t) = 2^{-\frac{j}{2}}\psi(-2^{-j}t), \quad (1)$$

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

где $h(t)$ – функция высокочастотного фильтра;

$g(t)$ – функция низкочастотного фильтра;

ψ, ϕ – скалирующие коэффициенты.

С каждым последующим применением фильтров к временному ряду отбрасывается все больше высокочастотных колебаний и в результате остается наиболее сглаженная компонента временного ряда. В целях прогнозирования исходного временного ряда, как правило, строится прогноз каждой компоненты. К примеру, в работе [32] для этих целей используются рекуррентные нейронные сети в стеке модели. После выполняется обратное вейвлет преобразование, которое приводит к исходной частности ряда. Описанная процедура в части сжатия данных с декомпозиций на частотные составляющие очень похожа на механизм сверточной нейронной сети. Существенное отличие кроется в структуре сверточных фильтров. В отличие от вейвлетов, в нейросети они не включают в себя априори заданные веса. Вместо этого веса определяются в процессе обучения. В результате фильтры сверточной сети выделяют не компоненты ряда с заданной частотой колебания, а компоненты ряда, необходимые последующим блокам нейронной сети для достижения сходимости целевой функции к заданной точности. Таким образом, в сверточной сети подчинение весов фильтров целевой функции обуславливает ее эффективность в обработке временных рядов и гарантирует выделение важных признаков для построения прогноза эндогенной переменной. Сеть LSTM использует отобранные признаки для поиска в них временных закономерностей и построения прогноза.

Нейронная сеть CNN-LSTM имеет множество гиперпараметров. К гиперпараметрам сверточной сети относятся: размерность временного окна входных данных (для обоих входов – высокочастотного и низкочастотного); количество сверточных фильтров; размер ядра; шаг фильтра; функции активации; параметры регуляризации. Для рекуррентной сети это количество рекуррентных ячеек на одном слое; количество рекуррентных слоёв функции активации; параметры регуляризации. Для сети в целом – это метод градиентного спуска, его параметры, размер подвыборки для одной итерации градиентного спуска (батч), число эпох обучения. Перечисленные параметры в значительной степени влияют на скорость обучения и производительность в зависимости от характеристик данных обучения.

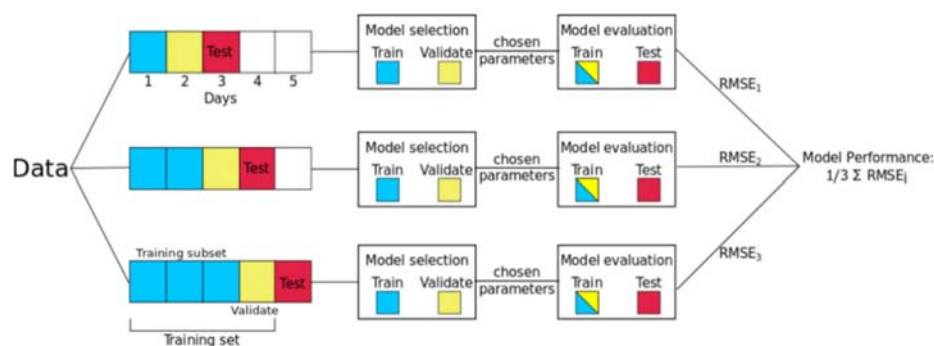
Для построения оптимальной архитектуры сети на данном этапе необходимо рассмотреть некоторые характеристики входных данных. В качестве прогнозируемой переменной выступает равновесная цена продажи электроэнергии. Среди объясняемых переменных имеются данные с месячной и почасовой частотой измерения. К месячным данным относятся четыре индекса промышленного производства. К почасовым данным: лаг цены, лаг равновесных объемов, среднесуточная температура и четыре специально сконструированных фактора для учета сезонных компонент объясняемой переменной. Прогноз строится на один час вперед. Одна прогнозная точка зависит от 72 лаговых значений почасовых данных

и трех лаговых значений месячных данных. Таким образом, три дня почасовых данных объясняют цену в следующий час. Одновременно с этим 3 точки, т.е. три предшествующих месяца объясняют каждое прогнозируемое значение следующего месяца. С переходом в эндогенной переменной к следующему месяцу окно сдвигается на месяц вперед. Таким образом, в случае с ИПП – динамика промышленного производства за 3 последних месяца обуславливает прогноз цены в следующем месяце в каждый час.

В результате шаблон данных для нейронной сети представляют две матрицы: с почасовыми данными размерности (72, 7) и месячными данными с размерностью (3, 4), а также соответствующее этим данным единственное прогнозируемое значение цены. В процессе исследования была разработана специальная оконная функция, разбивающая имеющийся пул данных на множество таких шаблонов.

Непосредственно перед оконным преобразованием и разделением данные приводятся к диапазону от 0 до 1 с применением минимаксного скалирования. Первый блок включает в себя два типа нейронных сетей: сверточную и полносвязную. Результатом его работы является конкатенация месячных и почасовые данных, которые приводятся к общей размерности. Почасовые данные, имеющие изначально размерность 72 по временной оси, пропускаются через сверточную сеть, которая состоит из 15 фильтров с ядром 2 и шагом 2. В качестве функции активации в каждом фильтре используется «Relu». В результате получается 15 векторов размерности 36. Месячные данные после входа также попадают в слой сверточной сети, однако в данном случае применяется 1-d сверка, которая снижает размерность до одного вектора, представляющего собой главную компоненту четырех макроэкономических факторов. Далее вектор попадает на вход в полносвязную нейронную сеть (FNN), которая состоит из одного слоя с 36 нейронами. Каждый нейрон использует линейную функцию активации. Выходной вектор добавляется в виде столбца к матрице 36*15, содержащей преобразованные почасовые данные. Расширенная матрица (36, 16) представляет собой промежуточную матрицу признаков, к которой также применяется 1-d сверка для выделения главной компоненты. Выделенная главная компонента передается в блок сверточной сети, который представлен двумя слоями. На первом находится 48 ячеек LSTM с функцией активации «Relu» и имеющих множественный выход. На втором слое находится 24 ячейки LSTM с единственным выходом и также функцией активации «Relu». На выходе из рекуррентного блока вектор размерности 24, который попадает в сеть внимания (Attention), а затем в единственный нейрон с линейной функцией активации. Выходом последнего нейрона является одно значение, которое и сравнивается с эталоном в процессе расчета ошибки отклонения.

Для обучения и тестирования нейронной выборка разделяется на обучающую, валидационную и тестовую. Отметим, что валидационная выборка не участвует напрямую в процессе обучения. Однако ошибка, вычисляемая на этих данных параллельно с ошибкой на тренировочной выборке по пришествию каждой эпохи, служит индикатором остановки обучения. Таким образом, валидационная выборка участвует в определении некоторых гиперпараметров нейронной сети, а именно: число эпох обучения и шаг градиентного спуска. Последний по некоторому правилу зависит от ошибки на валидационной выборке. К примеру, если ошибка по прошествии пяти эпох не сокращается, то шаг градиентного спуска сокращается в два раза. В результате для непредвзятой оценки качества прогноза необходима еще и тестовая выборка, которая не была никаким образом задействована при обучении. Стандартной практикой в задачах машинного обучения является кросс-валидация выборки, которая необходима для оценки обобщающей способности модели. Проверка эффективности модели на одном тестовом наборе данных может показать смещенную ошибку прогноза, поскольку нельзя с уверенностью утверждать, что независимые тестовые наборы не будут содержать иных паттернов поведения. Поскольку размерность данных позволяет, мы будем использовать стандартный способ кросс-валидации для временных рядов, идею которого иллюстрирует рис. 9.



Примечание – Источник: [33]

Рис. 9. Подход к кросс-валидации временных рядов

Вместо обучения и тестирования одной модели, согласно рисунку выше, необходимо обучить три независимые модели последовательно. Первая модель использует для обучения только 2/5 выборки (из них 1/5 для обучения, 1/5 для валидации) и следующую 1/5 выборки для расчета ошибки прогноза. Вторая модель использует 3/5 выборки для обучения и следующую 1/5 для теста. Третья модель обучается на 4/5 выборки и оставшуюся часть использует для расчета ошибки прогноза. Усреднение ошибок прогноза на тестовой выборке является конечной ошибкой модели. Выше был описан условный пример. В соб-

ственных эмпирических расчетах мы сформируем четыре выборки, последовательно отбрасывая один квартал от полного набора данных. Таким образом, как базовую модель, так и выбранные бенчмарки предстоит оценить четыре раза. И только после усреднения ошибок проводить сравнение подходов между собой.

В качестве оптимизатора для обучения нейронной сети используется «Adam». В процессе обучения нейронной сети на каждой эпохе рассчитывается среднеквадратическая ошибка отклонения на тестовой выборке и на валидационной. В том случае, если в течение трех эпох не происходит снижения ошибки, шаг градиентного спуска умножается на 0.2. Минимальным шагом может быть значение 0.001. Для одной итерации оптимизатора используется 256 шаблонов данных (батчей). Модель обучается на 30 эпохах, но останавливает обучение, если на протяжении 5-ти эпох ошибка на валидационной выборке не уменьшается.

Выше было приведено описание полной модели, однако в процессе тестирования предполагается блоки «FNN» и «Attention» включать последовательно. Это позволит оценить их влияние на процесс обучения и точность прогнозирования. В качестве бенчмарков выступают «ARIMA», классическая «LSTM» и модель «Midas». Каждая модель будет использоваться для прогнозирования итеративно. Однако в случае нейронных сетей также будет дополнительно протестирована архитектура, позволяющая одномоментно строить прогноз на 24 шага вперед. Для этого не требуется существенных конструктивных изменений в модели, однако необходима перестройка оконной функции для предобработки данных. Критериями для сравнения моделей могут быть: средняя квадратическая, средняя абсолютная и средняя абсолютная процентная ошибки, рассчитанные на прогнозируемом интервале без применения фактических данных в 24 часа и усредненные по кросс-валидационным множествам.

3. Моделирование цен на оптовом рынке электроэнергии (рынка на сутки вперед) и цен балансирующего рынка, в том числе с использованием данных атласа российской энергетики

3.1. Построение прогнозов цен на основе разработанного метода, сравнение результатов с другими методами прогнозирования временных рядов и совмещения размерности

В рамках данного раздела мы сосредоточимся на задаче моделирования цены на электричество на РСВ и декомпозиции цены на трендовую и сезонную составляющие. В рамках работы над моделью стохастической волатильности был определен ряд особенностей цен на рынке на сутки вперед и определены регрессоры, которые должны фигурировать в модели, чтобы эти особенности учитывать.

После выбора ряда интересующих нас регрессоров, необходимо определиться с функциональной формой модели: сравнить линейную модель с моделью с трансформацией и стохастической волатильностью.

Для начала перечислим регрессоры, которые будут фигурировать во всех моделях, которые будут сравниваться между собой с помощью критерия байесовского фактора.

— $y_t = \ln(S_t)$ – объясняемая переменная, логарифм средней за день цены в 1 ЦЗ и 2 ЦЗ,

— t – линейный тренд,

— $\sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t)$ – ряды Фурье 4 порядка для моделирования годовой сезонности,

— $d_{HDD,t}, d_{CDD,t}$ – дамми для отопительного и охлаждающего периода,

— $\Delta \ln(IPP_t)$ – разность логарифма трендовой составляющей ИПП,

— $\Delta \ln(p_{gas,t}), \Delta \ln(p_{coal,t})$ – разность логарифма среднемесячной цены по РФ на ресурсы (газ и уголь), для 1 ЦЗ цены на газ средние по 1 ЦЗ, цены на уголь – средние по РФ, для 2 ЦЗ цены на уголь средние по 2 ЦЗ, цены на газ – средние по РФ,

— $c_{Sat}, c_{Sun}, c_{Mon}, c_{hol}$ – дамми переменные на дни недели (суббота, воскресенье, понедельник) и праздничные дни.

Помимо этих факторов при дальнейшей разработке модели в качестве регрессоров могут быть также включены и другие показатели, прямо или косвенно отражающие влияние на цены на электричество:

- Расход электроэнергии на душу населения,
- Индекс промышленного производства по отраслям,
- Гидроэлектрический индекс производительности,
- Спотовая цена на Brent,
- Импорт сырой нефти на душу населения,
- Производство электроэнергии с помощью ВИЭ на душу населения,
- Баланс импорта-экспорта электроэнергии на душу населения.

Рассмотрим базовую модель – линейную регрессию без трансформации, построенную на основе модели [34].

Модель 0. Байесовская линейная регрессия, смотри уравнение (3):

$$\begin{aligned}
 y_t = & y_{t-1} + \mu + c_{trend} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD,t} + \\
 & + c_{CDD} \cdot d_{CDD,t} + c_{Sat} d_{Sat,t} + c_{Sun} d_{Sun,t} + c_{Mon} d_{Mon,t} + c_{hol} d_{hol,t} + \\
 & + c_{temppl} \cdot d_{\Delta_{temp} > 0, t} \cdot |\Delta_{temp, t}| + c_{tempn} \cdot d_{\Delta_{temp} < 0, t} \cdot |\Delta_{temp, t}| + \\
 & + c_{temppl} \cdot d_{\Delta_{temp} > 0, t-1} \cdot |\Delta_{temp, t-1}| + c_{tempnl} \cdot d_{\Delta_{temp} < 0, t-1} \cdot |\Delta_{temp, t-1}| + \\
 & + c_{IPP} \cdot \Delta \ln(IPP_t) + c_{gas} \cdot \Delta \ln(p_{gas, t}) + c_{coal} \cdot \Delta \ln(p_{coal, t}) + \epsilon_t
 \end{aligned} \tag{3}$$

где $\epsilon_t \sim N(0, \sigma)$, $\sigma \sim \text{Cauchy}(0, 5) * \mathbb{I}(0, \infty)$.

Такая одноуровневая спецификация модели аналогична модели, представленной в [34] с добавлением дополнительных регрессоров, выбранных в процессе работы над моделью стохастической волатильности. Однако несмотря на то, что одноуровневая спецификация позволяет получить преимущества при прогнозировании по сравнению с моделями ARIMA и различными схемами усреднения, необходимо понимать, что она не позволяет учесть факторы сезонности и трендовую составляющую в полной мере, поскольку моделируются именно приросты цен (коэффициент при лаге равен единице).

В предыдущей работе было предложено использовать трансформацию переменных. Благодаря значительной гибкости при построении моделей с использованием байесовского подхода в рамках одной модели можно использовать двухуровневую структуру, где на одном уровне учитываются все факторы, связанные с трендом и сезонностью, а на другом моделируется уже очищенный от тренда и сезонности ряд по модели, предложенной в статье [34].

Модель 1 Байесовская линейная регрессия, смотри уравнения (4) и (5):

$$\begin{aligned}
 r_t = & p_t - \left(\mu + c_{trend} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD,t} + \right. \\
 & \left. + c_{CDD} \cdot d_{CDD,t} + c_{Sat} d_{Sat,t} + c_{Sun} d_{Sun,t} + c_{Mon} d_{Mon,t} + c_{hol} d_{hol,t} \right)
 \end{aligned} \tag{4}$$

$$\begin{aligned}
r_t = & r_{t-1} + c_{tempp} \cdot d_{\Delta_{temp}>0,t} \cdot |\Delta_{temp,t}| + c_{tempn} \cdot d_{\Delta_{temp}<0,t} \cdot |\Delta_{temp,t}| + \\
& + c_{temppl} \cdot d_{\Delta_{temp}>0,t-1} \cdot |\Delta_{temp,t-1}| + c_{tempnl} \cdot d_{\Delta_{temp}<0,t-1} \cdot |\Delta_{temp,t-1}| + \\
& + c_{IPP} \cdot \Delta \ln(IPP_t) + c_{gas} \cdot \Delta \ln(p_{gas,t}) + c_{coal} \cdot \Delta \ln(p_{coal,t}) + \epsilon_t
\end{aligned} \tag{5}$$

где r_t – очищенный от сезонности и тренда ряд,
 $\epsilon_t \sim N(0, \sigma)$, $\sigma \sim \text{Cauchy}(0, 5) * \mathbb{I}(0, \infty)$.

Сравнение двух моделей по критерию байесовского фактора указывает на то, что данные на 100% более вероятны в рамках Модели 1 по сравнению с Моделью 0, как для 1 ЦЗ, так и для 2 ЦЗ. Таким образом, при дальнейшем моделировании цен на электричество будет использоваться только модели с трансформацией.

Следующим шагом усложнения модели является добавление стохастической волатильности. В литературе по моделированию и прогнозированию цен на электричество предполагается, что волатильность цен на электричество не только не является постоянным параметром модели, но изменяется согласно определённом случайному процессу. В модели [34] также учитывается эффект рычага, благодаря которому можно учесть, что большим значениям логарифма цены соответствуют большие значения дисперсии. На данном этапе мы рассматриваем базовую спецификацию модели стохастической волатильности, которую в дальнейшем можно будет усложнить, при условии, что критерий байесовского фактора покажет, что данные более вероятны в рамках модели стохастической волатильности по сравнению с моделью, где волатильности цен задается фиксированным параметром.

Модель 2. Модель стохастической волатильности с трансформацией, смотри уравнения (6)-(8):

$$\begin{aligned}
r_t = & p_t - \left(\mu + c_{trend} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD,t} + \right. \\
& \left. + c_{CDD} \cdot d_{CDD,t} + c_{Sat} d_{Sat,t} + c_{Sun} d_{Sun,t} + c_{Mon} d_{Mon,t} + c_{hol} d_{hol,t} \right)
\end{aligned} \tag{6}$$

$$\begin{aligned}
r_t = & r_{t-1} + c_{tempp} \cdot d_{\Delta_{temp}>0,t} \cdot |\Delta_{temp,t}| + c_{tempn} \cdot d_{\Delta_{temp}<0,t} \cdot |\Delta_{temp,t}| + \\
& + c_{temppl} \cdot d_{\Delta_{temp}>0,t-1} \cdot |\Delta_{temp,t-1}| + c_{tempnl} \cdot d_{\Delta_{temp}<0,t-1} \cdot |\Delta_{temp,t-1}| + \\
& + c_{IPP} \cdot \Delta \ln(IPP_t) + c_{gas} \cdot \Delta \ln(p_{gas,t}) + c_{coal} \cdot \Delta \ln(p_{coal,t}) + \epsilon_t^{(1)} \cdot \exp(h_t/2)
\end{aligned} \tag{7}$$

$$h_t = \kappa + \phi(h_{t-1} - \kappa) + \epsilon_t^{(2)} \sigma \tag{8}$$

где h_t – логарифм волатильности,

κ – среднее логарифма волатильности или долгосрочная волатильность,

ϕ – коэффициент устойчивости волатильности или частота, с которой логарифм волатильности возвращается к долгосрочному среднему,

$$\varepsilon_t, \epsilon_t^{(1)}, \epsilon_t^{(2)} \sim N(0, \sigma) \text{ i.i.d.}$$

Сравнение Модели 1 и Модели 2 по критерию байесовского фактора указывает на то, что данные на 100% более вероятны в рамках Модели 2 по сравнению с Моделью 1, как для 1 ЦЗ, так и для 2 ЦЗ. Таким образом, при дальнейшем моделировании цен на электричество будет использоваться только модели стохастической волатильности.

В моделях 0–2 использовались неинформативные априорные распределения коэффициентов, смотри уравнения (9):

$$\begin{aligned} \mu &\sim N(7, 1), \alpha_i, \beta_i, c_{trend}, c_{Sat}, c_{Sun}, c_{Mon}, c_{hol}, c_{CDD}, c_{HDD} \sim N(0, 1) \\ c_{tempp}, c_{temppl}, c_{tempn}, c_{temnl}, c_{IPP, gas}, c_{coal} &\sim Laplace(0, 1); \\ \sigma &\sim Cauchy(0, 5) \cdot \mathbb{I}_{(0, \infty)}, \phi \sim U(-1, 1); \kappa \sim Cauchy(0, 10) \end{aligned} \quad (9)$$

Ниже в таблицах 2-3, позволяющих сравнить оценки коэффициентов, полученных по Моделям 0 – 4, для 1 и 2 ЦЗ, видно, что низкочастотные регрессоры (цены на ресурсы и ИПП) не оказывают значимого влияния на цены, однако из литературы по моделированию цен на электричество известно, что эти факторы должны иметь значимое положительное влияние. Для того чтобы проверить эту гипотезу мы используем информативные априорные распределения при той же спецификации модели.

Модель 3. Модель стохастической волатильности с трансформацией с информативными априорными распределениями, смотри уравнения (10)-(12):

$$\begin{aligned} r_t = p_t - &\left(\mu + c_{trend} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD, t} + \right. \\ &\left. + c_{CDD} \cdot d_{CDD, t} + c_{Sat} d_{Sat, t} + c_{Sun} d_{Sun, t} + c_{Mon} d_{Mon, t} + c_{hol} d_{hol, t} \right) \end{aligned} \quad (10)$$

$$\begin{aligned} r_t = r_{t-1} + &c_{tempp} \cdot d_{\Delta_{temp} > 0, t} \cdot |\Delta_{temp, t}| + c_{tempn} \cdot d_{\Delta_{temp} < 0, t} \cdot |\Delta_{temp, t}| + \\ &+ c_{temppl} \cdot d_{\Delta_{temp} > 0, t-1} \cdot |\Delta_{temp, t-1}| + c_{temnl} \cdot d_{\Delta_{temp} < 0, t-1} \cdot |\Delta_{temp, t-1}| + \\ &+ c_{IPP} \cdot \Delta \ln(IPP_t) + c_{gas} \cdot \Delta \ln(p_{gas, t}) + c_{coal} \cdot \Delta \ln(p_{coal, t}) + \epsilon_t^{(1)} \cdot \exp(h_t/2) \end{aligned} \quad (11)$$

$$h_t = \kappa + \phi(h_{t-1} - \kappa) + \epsilon_{t-1}^{(2)} \sigma \quad (12)$$

где h_t – логарифм волатильности,
 κ – среднее логарифма волатильности или долгосрочная волатильность,
 ϕ – коэффициент устойчивости волатильности или частота, с которой логарифм волатильности возвращается к долгосрочному среднему,

$$\varepsilon_t, \epsilon_t^{(1)}, \epsilon_t^{(2)} \sim N(0, \sigma) \text{ i.i.d.}$$

Априорные распределения коэффициентов представлены в уравнение (13):

$$\begin{aligned} \mu &\sim N(7, 1), \alpha_i, \beta_i, c_{trend}, c_{Sat}, c_{Sun}, c_{Mon}, c_{hol}, c_{CDD}, c_{HDD} \sim N(0, 1) ; \\ c_{trend}, c_{CDD}, c_{HDD} &\sim N(0, 1) \cdot \mathbb{I}_{(0, \infty)} \\ c_{tempp}, c_{temmpl}, c_{tempn}, c_{temnl} &\sim Laplace(0, 1) \\ c_{IPP}, c_{gas}, c_{coal} &\sim Laplace(0, 1) \cdot \mathbb{I}_{(0, \infty)}; \\ \sigma &\sim Cauchy(0, 5) \cdot \mathbb{I}_{(0, \infty)}, \phi \sim U(-1, 1); \kappa \sim Cauchy(0, 10) \end{aligned} \quad (13)$$

Сравнение Модели 2 и Модели 3 по критерию байесовского фактора указывает на то, что данные для 1 ЦЗ на 100% более вероятны в рамках Модели 2, не предполагающей априорно никакого эффекта ценовых факторов, по сравнению с Моделью 3, в то время как для 2 ЦЗ данные 99,9% более вероятны в рамках Модели 3, предполагающей положительный эффект отопительного/охладительного периода, ИПП и цен на ресурсы, чем в модели, которая не предполагает никакого эффекта. Таким образом, можно сделать вывод о том, что даже несмотря на то, что коэффициенты при низкочастотных регрессорах незначимы, по крайней мере для 2 ЦЗ гипотеза об одновременном положительном эффекте всех этих факторов не отвергается.

Такой вывод позволяет говорить о том, что необходимо продолжать разрабатывать модель и корректировать спецификацию модели таким образом, чтобы можно было учесть низкочастотные регрессоры, поскольку на данном этапе включение низкочастотных регрессоров с использованием простой интерполяции не позволяет уловить значимое влияние неценовых факторов по отдельности.

Следующим шагом эмпирической стратегии является проверка гипотезы о кусочно-линейном тренде. На данный момент ни одна модель не позволяет получить значимые оценки коэффициента при линейном тренде. В предыдущих работах при анализе с помощью модели Prophet было показано, что вероятно, трендовая составляющая ряда имеет кусочно-линейный характер.

Сравним Модель 2 и Модель 4, где мы учитываем кусочно-линейный характер трендовой составляющей с помощью набора дамми переменных, разделяющих ряд на 4

блока для 1 ЦЗ и на 3 блока для 2 ЦЗ, причем точки перелома используются те же, что и были определены моделью Prophet.

Модель 4. Модель стохастической волатильности с кусочно-линейным трендом представлена уравнениями (14)-(16):

$$r_t = p_t - \left(\mu + c_{trend} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD,t} + \right. \\ \left. + c_{CDD} \cdot d_{CDD,t} + c_{Sat} d_{Sat,t} + c_{Sun} d_{Sun,t} + c_{Mon} d_{Mon,t} + c_{hol} d_{hol,t} \right) \quad (14)$$

$$r_t = r_{t-1} + c_{tempp} \cdot d_{\Delta_{temp}>0,t} \cdot |\Delta_{temp,t}| + c_{tempn} \cdot d_{\Delta_{temp}<0,t} \cdot |\Delta_{temp,t}| + \\ + c_{temppl} \cdot d_{\Delta_{temp}>0,t-1} \cdot |\Delta_{temp,t-1}| + c_{tempnl} \cdot d_{\Delta_{temp}<0,t-1} \cdot |\Delta_{temp,t-1}| + \\ + c_{IPP} \cdot \Delta \ln(IPP_t) + c_{gas} \cdot \Delta \ln(p_{gas,t}) + c_{coal} \cdot \Delta \ln(p_{coal,t}) + \epsilon_t^{(1)} \cdot \exp(h_t/2) \quad (15)$$

$$h_t = \kappa + \phi(h_{t-1} - \kappa) + \epsilon_{t-1}^{(2)} \sigma \quad (16)$$

где h_t – логарифм волатильности,

κ – среднее логарифма волатильности или долгосрочная волатильность,

ϕ – коэффициент устойчивости волатильности или частота, с которой логарифм

волатильности возвращается к долгосрочному среднему, $\epsilon_t, \epsilon_t^{(1)}, \epsilon_t^{(2)} \sim N(0, \sigma)$ i.i.d.

Сравнение двух моделей по критерию байесовского фактора указывает на то, что данные на 100% более вероятны в рамках Модели 2 по сравнению с Моделью 4, как для 1 ЦЗ, так и для 2 ЦЗ. Таким образом, можно сделать вывод, что точки перелома тренда моделью Prophet были определены неверно. Причиной этому может быть тот факт, что модель не учитывала стохастическую волатильность ряда, а также влияние ценовых факторов на разных уровнях.

Рассмотрим подробнее оценки, полученные по Моделям 0-4 для 1 ЦЗ в таблице 2, для 2 ЦЗ в таблице 3.

Таблица 2

Оценки по Моделям 0–4 для 1 ценовой зоны

Регрес- сор	Характе- ристика	Модель 0	Модель 1	Модель 2	Модель 3	Модель 4
mu						(7.06319;6.93196;7.03798;6.70996)
	Оценка	0.02812	7.03139	7.07194	7.11719	
	N eff	89.0	6495.0	15416.0	14604.0	(681.0;1723.0;1882.0;1928.0)
	s.d.	0.01769	0.16915	0.13616	0.12507	(0.13822;0.77615;0.90877;0.95795)
			*	*	*	*
ctrend						(-0.00082;-0.0011;-0.00123;-0.00104)
	Оценка	-0,00001	-0.0003	-0.00099	0.00094	
	N eff	1162.0	15099.0	11048.0	10275.0	(729.0;963.0;952.0;941.0)
	s.d.	0,00001	0.00217	0.00164	0.00077	(0.00186;0.00178;0.00172;0.00168)
					(*)	
ccdd	Оценка	0.00921	-0.00991	-0.00882	0.00326	-0.00887
	N eff	85.0	13482.0	10600.0	10129.0	858.0
	s.d.	0.02089	0.00767	0.00697	0.00288	0.00692
					*	
ccoal	Оценка	-0.00075	0.00029	-0.00002	0.00072	-0.00004
	N eff	124.0	10647.0	12548.0	9812.0	2590.0
	s.d.	0.00436	0.00124	0.00099	0.00056	0.00097
					*	
cgas	Оценка	-0.00021	-0.00008	-0.00003	0.00006	-0.00003
	N eff	465.0	9202.0	10475.0	9635.0	2319.0
	s.d.	0.00048	0.00013	0.00009	0.00005	0.00009
cgdp	Оценка	0.28028	0.0097	0.00565	0.98709	-0.03427
	N eff	2.0	9742.0	16372.0	11879.0	1441.0
	s.d.	1.35497	1.46248	1.40941	0.9657	1.35823
					*	
chdd	Оценка	-0.00795	0.01831	0.00631	0.01616	0.00607
	N eff	78.0	13528.0	13474.0	8165.0	1580.0
	s.d.	0.02534	0.01551	0.01604	0.01135	0.01582
					*	
chol	Оценка	-0.02013	-0.03703	-0.02881	-0.02845	-0.02805
	N eff	90.0	13888.0	9196.0	10349.0	1003.0
	s.d.	0.01702	0.00581	0.00618	0.00623	0.00615
			*	*	*	*
cmon	Оценка	0.06265	0.01554	0.01186	0.01202	0.0119
	N eff	89.0	12126.0	11188.0	11104.0	1077.0
	s.d.	0.01278	0.00259	0.00203	0.00201	0.002
		*	*	*	*	*
csat	Оценка	-0.02573	-0.01225	-0.01261	-0.0126	-0.01259
	N eff	74.0	11839.0	10499.0	9910.0	1269.0

Продолжение таблицы 2

Регрес- сор	Характе- ристика	Модель 0	Модель 1	Модель 2	Модель 3	Модель 4
	s.d.	0.01249	0.00267	0.00213	0.00216	0.00215
		*	*	*	*	*
csun	Оценка	-0.0667	-0.06374	-0.06102	-0.06085	-0.06099
	N eff	156.0	10584.0	10234.0	10972.0	870.0
	s.d.	0.01216	0.00309	0.00248	0.00247	0.00251
		*	*	*	*	*
ctempn	Оценка	0.00546	0.00427	0.00333	0.00349	0.00334
	N eff	97.0	13012.0	9619.0	9017.0	1301.0
	s.d.	0.00581	0.00165	0.00121	0.00119	0.00119
			*	*	*	*
ctempnl	Оценка	-0.00496	-0.00476	-0.00408	-0.00383	-0.00411
	N eff	95.0	13422.0	9342.0	9370.0	1332.0
	s.d.	0.00586	0.00166	0.00121	0.00119	0.00119
			*	*	*	*
ctemp	Оценка	0.00091	0.00242	0.00436	0.00378	0.00443
	N eff	24.0	11793.0	13678.0	10207.0	1299.0
	s.d.	0.00698	0.00199	0.0017	0.00167	0.0017
				*	*	*
ctemppl	Оценка	-0.00335	-0.00266	-0.00428	-0.0042	-0.00432
	N eff	24.0	12396.0	13324.0	10532.0	1284.0
	s.d.	0.00677	0.00197	0.00172	0.00167	0.00171
				*	*	*
sigma	Оценка	0.17204	0.04852			
	N eff	125.0	12503.0			
	s.d.	0.00263	0.00079			
		*	*			
kap	Оценка			-6.47658	-6.48057	-6.40225
	N eff			3778.0	5099.0	5321.0
	s.d.			0.08284	0.08041	0.08141
				*	*	*
phi	Оценка			0.8507	0.84208	0.84393
	N eff			134.0	142.0	135.0
	s.d.			0.03298	0.03113	0.03283
				*	*	*
log-posterior	Оценка	2361.8976	4795.56079	5741.83328	5642.37826	5792.0171

Примечание – Источник: составлено авторами.

Таблица 3

Оценки по Моделям 0–4 для 2 ценовой зоны

Регрес- сор	Харак- тери- стика	Модель 0	Модель 1	Модель 2	Модель 3	Модель 4
mu	Оценка	0.02884	6.46217	6.47706	6.49975	(6.46338;6.97846;7.07976)
	N eff	119.0	8503.0	8997.0	11507.0	(6857.0;16343.0;18325.0)
	s.d.	0.01804	0.45107	0.22156	0.20919	(0.22284;0.8897;0.9476)
			*	*	*	*
ctrend	Оценка	-0,00001	0.00056	-0.00328	0.00099	(-0.002;-0.00348;-0.00329)
	N eff	821.0	12575.0	10621.0	12228.0	(8117.0;8471.0;8340.0)
	s.d.	0,00001	0.00678	0.00225	0.00087	(0.00256;0.00242;0.00226)
					(*)	
ccdd	Оценка	-0.05477	0.01528	-0.01257	0.00793	-0.01267
	N eff	128.0	14005.0	10199.0	9868.0	8518.0
	s.d.	0.02615	0.02368	0.01415	0.00692	0.0142
		*			*	
ccoal	Оценка	0.00815	0.0011	0.00176	0.0024	0.00181
	N eff	148.0	11952.0	10591.0	10013.0	10097.0
	s.d.	0.00854	0.00658	0.00235	0.00169	0.00238
					*	
cgas	Оценка	-0.00052	0.0002	-0.00027	0.00019	-0.0003
	N eff	255.0	12139.0	10103.0	11412.0	10343.0
	s.d.	0.00152	0.00104	0.00034	0.00016	0.00034
					*	
cgdp	Оценка	0.25604	-0.02619	-0.00152	0.98989	0.01092
	N eff	2.0	9404.0	17729.0	12150.0	12146.0
	s.d.	0.9249	1.42435	1.41243	0.97968	1.47129
					*	
chdd	Оценка	-0.01187	0.00053	0.01503	0.02229	0.01531
	N eff	84.0	16781.0	11492.0	7857.0	10302.0
	s.d.	0.02779	0.05017	0.01799	0.01345	0.01817
					*	
chol	Оценка	-0.00331	0.03311	0.00897	0.00958	0.00974
	N eff	99.0	14048.0	8549.0	9322.0	6670.0
	s.d.	0.01925	0.01917	0.00716	0.00717	0.00723
cmon	Оценка	-0.00042	0.02569	0.00611	0.00654	0.0062
	N eff	134.0	10423.0	9409.0	11837.0	8049.0
	s.d.	0.01612	0.00873	0.00296	0.00302	0.00299
			*	*	*	*
csat	Оценка	0.01231	0.02747	0.0148	0.01472	0.01495
	N eff	155.0	11185.0	8897.0	10150.0	6363.0
	s.d.	0.01551	0.00878	0.00315	0.00313	0.00312
			*	*	*	*
csun	Оценка	-0.03352	0.0081	-0.01197	-0.01208	-0.01185

Продолжение таблицы 3

Регрес- сор	Харак- тери- стика	Модель 0	Модель 1	Модель 2	Модель 3	Модель 4
	N eff	90.0	8191.0	8949.0	9001.0	6383.0
	s.d.	0.01653	0.01016	0.00376	0.00371	0.00376
		*		*	*	*
ctempn	Оценка	-0.00192	-0.00196	0.0015	0.00191	0.00161
	N eff	129.0	10400.0	10556.0	10552.0	9206.0
	s.d.	0.00489	0.00357	0.00126	0.00124	0.00127
ctempnl	Оценка	0.00178	0.00116	-0.00263	-0.00243	-0.00268
	N eff	120.0	10758.0	10593.0	10167.0	9111.0
	s.d.	0.00482	0.00359	0.00125	0.00122	0.00125
				*		*
ctempp	Оценка	-0.01037	-0.01119	-0.00353	-0.0043	-0.00352
	N eff	88.0	9960.0	7915.0	8680.0	7248.0
	s.d.	0.00546	0.00408	0.00136	0.00137	0.00137
			*	*	*	*
ctemppl	Оценка	0.00983	0.01012	0.00362	0.00357	0.00353
	N eff	90.0	9976.0	8322.0	8777.0	7336.0
	s.d.	0.00538	0.00407	0.00133	0.00134	0.00133
			*	*	*	*
sigma	Оценка	0.21702	0.15899			
	N eff	171.0	15237.0			
	s.d.	0.00336	0.00262			
		*	*			
kap	Оценка			-5.39896	-5.38897	-5.40225
	N eff			8411.0	10862.0	12321.0
	s.d.			0.12984	0.12909	0.12978
				*	*	*
phi	Оценка			0.8448	0.84842	0.84393
	N eff			473.0	283.0	375.0
	s.d.			0.02398	0.02354	0.02283
				*	*	*
log-posterior	Оценка	1924.59971	2541.59265	3618.2979	3613.15596	3612.91731

Примечание – Источник: составлено авторами.

По результатам оценивания можно отметить, что в 1 ЦЗ наблюдается статистически значимое негативное влияние праздничных дней на цену (снижение примерно на 3%), во 2 ЦЗ влияния праздников нет.

Также отмечается наличие недельной сезонности: в понедельник цены выше, чем в остальные дни, причем этот эффект более заметен в 1 ЦЗ (на 1% в 1 ЦЗ, на 0.6 % во 2 ЦЗ), в воскресенье цены ниже, чем в остальные дни, причем этот эффект более заметен в 1 ЦЗ (на 6% в 1 ЦЗ, на 1% во 2 ЦЗ), в субботу в 1 ЦЗ цены ниже на 1,2%, чем в остальные дни, во 2 ЦЗ цены выше на 1,4%, чем в остальные дни.

Влияние дамми-переменной отопительного и охлаждающего периода, вероятно, было нивелировано включением в модель сезонности с помощью рядов Фурье, во всех моделях с трансформацией было обнаружено влияние отклонения температуры от

плавающего среднего, причем в 1 ЦЗ сегодняшние отклонения от ожидаемой температуры, как вверх, так и вниз положительно влияют на цену, вчерашние отклонения от ожидаемой температуры, как вверх, так и вниз отрицательно влияют на цену, во 2 ЦЗ обнаружено только отрицательное влияние сегодняшнего отклонения температуры вверх, отрицательное (либо никакого, если по BF критерию выбирается Модель с информативными априорными) влияние вчерашнего отклонения температуры вниз и положительное влияние вчерашнего отклонения температуры вверх.

Следующим этапом доработки модели будет добавление скачкообразной компоненты и эффекта рычага, использование кусочно-заданной линейной функции для тренда как в модели Prophet, а также работа с разночастотными данными.

Перейдем к пятому этапу построения модели - добавлению в модель скачкообразной компоненты. В статье [34] скачки моделировались как ненаблюдаемая компонента – случайная величина, которая задается как в уравнении (17) ниже:

$$q_t = \begin{cases} 1, & p_D \\ 0, & p_0 \\ -1, & p_U \end{cases} \quad (17)$$

где величины скачков вверх и вниз распределены экспоненциально, с разными параметрами $\xi_f^U \sim \text{Exp}(\eta_U)$ i.i.d., $\xi_f^D \sim \text{Exp}(\eta_D)$ i.i.d.

Для того, чтобы оценить, влияет ли добавление скачков в модель на стандартные ошибки оценок коэффициентов, добавим скачки в Модель 2 (модель стохастической волатильности с трансформацией), поскольку было выявлено, что именно эта модель описывает данные наилучшим образом (по критерию байесовского фактора). Ниже представлено уравнение (18) для Модели 5 (модели со скачками):

$$\begin{aligned} r_t &= p_t - \left(\mu + c_{\text{trend}} \cdot t + \sum_{i=1}^4 (\alpha_i \cos 2\pi f_i t + \beta_i \sin 2\pi f_i t) + c_{HDD} \cdot d_{HDD,t} + \right. \\ &\quad \left. + c_{CDD} \cdot d_{CDD,t} + c_{Sat} d_{Sat,t} + c_{Sun} d_{Sun,t} + c_{Mon} d_{Mon,t} + c_{hol} d_{hol,t} + J_t \right) \\ r_t &= r_{t-1} + c_{\text{temp}} \cdot d_{\Delta_{\text{temp}} > 0, t} \cdot |\Delta_{\text{temp}, t}| + c_{\text{tempn}} \cdot d_{\Delta_{\text{temp}} < 0, t} \cdot |\Delta_{\text{temp}, t}| + \\ &\quad + c_{\text{temppl}} \cdot d_{\Delta_{\text{temp}} > 0, t-1} \cdot |\Delta_{\text{temp}, t-1}| + c_{\text{tempnl}} \cdot d_{\Delta_{\text{temp}} < 0, t-1} \cdot |\Delta_{\text{temp}, t-1}| + \\ &\quad + c_{IPP} \cdot \Delta \ln(IPP_t) + c_{\text{gas}} \cdot \Delta \ln(p_{\text{gas}, t}) + c_{\text{coal}} \cdot \Delta \ln(p_{\text{coal}, t}) + \epsilon_t \\ J_t &= -\xi_t^D \cdot \mathbb{I}(q_t = -1) + 0 \cdot \mathbb{I}(q_t = 0) + \xi_t^U \cdot \mathbb{I}(q_t = 1) \end{aligned} \quad (18)$$

где q_t – ненаблюдаемый вектор наличия скачка вверх или вниз, дискретная случайная величина, которая может принимать значения $\{-1, 0, 1\}$,

η_u – ненаблюдаемая экспоненциально распределенная величина скачка вверх с параметром ν_u ,

η_d – ненаблюдаемая экспоненциально распределенная величина скачка вверх с параметром ν_d .

В языке программирования Stan для байесовского моделирования для того, чтобы можно было моделировать дискретные параметры (случайные величины) используется модели смесей (mixture models).

В моделях конечных смесей предполагается существование нескольких механизмов, порождающих данные. В явном виде модель выглядит так: сначала генерируется индекс, отражающий то, по какой модели будут генерироваться данные. Каждое состояние имеет вероятность θ_k , причем $\sum_k \theta_k = 1$, смотри уравнение (19):

$$z_i \sim \text{categorical}(\theta) \quad (19)$$

Далее после выбора модели определяется выбирается распределение для генерирования наблюдения. Например, если рассматривается модель смеси нормальных распределений, то выбирается соответствующее среднее μ_k и параметр масштаба σ_k , смотри уравнение (20):

$$y_i \sim \text{normal}(\mu_{z_i}, \sigma_{z_i}) \quad (20)$$

Модель для генерации y_i не обязательно должна быть такой простой, и, конечно, не ограничивается нормальными смесями. Например, в случае модели со скачками вместо y_i будет генерироваться переменная J_t .

Вклад в правдоподобие от наблюдения y_i зависит от генеративной плотности распределения y_i при заданном наборе параметров. Если вероятность, что наблюдаемый y_i был взят из модели 1 равна θ_1 , из модели 2 – θ_2 , и так далее, то условная вероятность наблюдения y_i при заданных параметрах задается уравнением (21):

$$p(y_i | \theta, \mu, \sigma) = \sum_{k=1}^3 \theta_k \times \text{normal}(y_i | \mu_k, \sigma_k) \quad (21)$$

где $\text{normal}(y_i | \mu_k, \sigma_k)$ – плотность нормального распределения в точке y_i , отражающая вероятностный вклад в правдоподобие модели от одной точки данных.

Вероятность всей модели, в предположении, что наблюдения независимы, является произведением индивидуальных вкладов в правдоподобие. При логарифмировании получаем уравнение (22):

$$\log(p(y_i|\Theta, \mu, \sigma)) = \log\left(\sum_{k=1}^3 \theta_k \times \text{normal}(y_i|\mu_k, \sigma_k)\right) \quad (22)$$

Функция правдоподобия может быть переписана в следующем виде, смотри уравнение (23):

$$\log(p(y_i|\Theta, \mu, \sigma)) = \log\left(\sum_{k=1}^3 \exp\left(\log(\theta_k) + \log(\text{normal}(y_i|\mu_k, \sigma_k))\right)\right) \quad (23)$$

Это значительно облегчает вычисления при реализации модели конечной смеси в Stan.

Для модели 5 необходимо моделировать J_t , как модель смеси. В таком случае вероятностный вклад в правдоподобие от компоненты скачка, выраженной уравнением (24):

$$p(J_i|\Theta, \mu, \sigma) = p_d \times \text{NegExponential}(y_i|v_d) + p_0 + p_u \times \text{Exponential}(y_i|v_u) \quad (24)$$

где $\text{Exponential}(y_i|v_u)$ – плотность экспоненциального распределения,

$\text{NegExponential}(y_i|v_d)$ – плотность отрицательного экспоненциального распределения, функция плотности для которого задается уравнением (25):

$$f(x|\varepsilon, \lambda) = \begin{cases} 0 & x \geq \varepsilon \\ \lambda e^{-\lambda(\varepsilon-x)} & x < \varepsilon \end{cases} \quad (25)$$

где для стандартного экспоненциального распределения $\varepsilon = 0$, соответственно, в уравнении выше этот параметр опускается, а отрицательное экспоненциальное распределение отличается от стандартного тем, что носитель функции – отрицательная часть оси $(-\infty, 0]$.

Таким образом, в модели будет фигурировать 3 дополнительные ненаблюдаемые переменные: первая наличие скачка, уравнение (26):

$$q_t = \begin{cases} 1, & p_d \\ 0, & p_0 \\ -1, & p_u \end{cases} \quad (26)$$

где $q_t \sim \text{Categorical}(\{1,2,3\}; \Theta) - 2$, $\Theta = \{p_d, p_0, p_u\}$. Поскольку $p_d + p_0 + p_u = 1$, в Stan вектор $[p_d, p_0, p_u]$ задается как симплекс.

Априорное распределение для вероятностей скачка – распределение Дирихле, уравнение (27):

$$\Theta \sim \text{Dirichlet}(2,2,2) \quad (27)$$

Ненаблюдаемые величины скачков вверх и вниз соответственно, уравнение (28):

$$\xi_f^U \sim \text{Exp}(\eta_U) i.i.d., \xi_f^D \sim \text{Exp}(\eta_D) i.i.d., \quad (28)$$

где априорное распределение для параметров экспоненциального распределения задается как обратное гамма распределение с теми же параметрами, что и в модели, представленной уравнением (29):

$$\eta_U \sim \text{IG}(1.86, 0.43) i.i.d., \eta_D \sim \text{IG}(1.86, 0.43) i.i.d. \quad (29)$$

Ниже представлены оценки по Модели 5 (модели стохастической волатильности с информативными априорными) для 1 в таблице 4 и 2 ценовой зоны в таблице 5.

Основным результатом модели 5 является то, что из-за включения трех дополнительных ненаблюдаемых переменных число, оцениваемых параметров увеличилось в 4 раза (учитывая ненаблюдаемую переменную волатильности, для которой оценивается 988 параметров, т.е. по параметру в каждой точке). Также по критерию байесовского фактора модели без скачков на 100% более вероятны, чем модели со скачками (для обеих ценовых зон).

Также при диагностике сходимости MCMC по распределению по статистике Rhat, можно отметить, что для некоторых коэффициентов Rhat отклоняется от 1 более, чем на 50%, что означает, что цепи не очень хорошо смешались, отражая проблемы в спецификации самой модели. Таким образом критерии надежности байесовских оценок отражают тот факт, что модели со скачками в спецификации [34] не подходят для анализа данных временных рядов.

Таблица 4

Оценки по модели 5 для 1 ЦЗ.

Параметр	n_{eff}	mean	sd	2.5%	50%	97.5%
ccdd	23	0.012	0.013	0.000	0.005	0.041
chdd	32	0.049	0.028	0.014	0.039	0.102
ctempp	51	0.003	0.002	0.000	0.003	0.008
ctempn	35	0.007	0.003	0.003	0.007	0.013
ctemppl	32	-0.004	0.003	-0.010	-0.003	-0.001
ctempnl	27	-0.006	0.002	-0.010	-0.006	-0.003
csat	33	-0.011	0.005	-0.017	-0.011	0.001
csun	42	-0.061	0.006	-0.073	-0.060	-0.053
emon	111	0.013	0.004	0.006	0.013	0.023
chol	23	-0.027	0.017	-0.055	-0.032	-0.001
ctrend	12	0.004	0.004	0.000	0.002	0.012
cipp	28	0.900	0.927	0.043	0.647	3.723
cgas	13	0.001	0.001	0.000	0.001	0.001
ccoal	13	0.002	0.001	0.000	0.001	0.005
mu	25	7.204	0.207	6.715	7.253	7.488
kap	2	-4.823	1.433	-6.370	-5.529	-2.405
phi	3	0.109	0.123	-0.068	0.059	0.346
nu_d	26	245.187	99.977	4.281	34.669	1126.28
nu_u	35	0.375	0.288	0.115	0.270	1.131

Примечание – значимые на 5% уровне коэффициенты с априорными распределениями на всей оси $(-\infty, \infty)$ и со средним в 0 отмечены жирным шрифтом

Таблица 5

Оценки по модели 5 для 2 ЦЗ

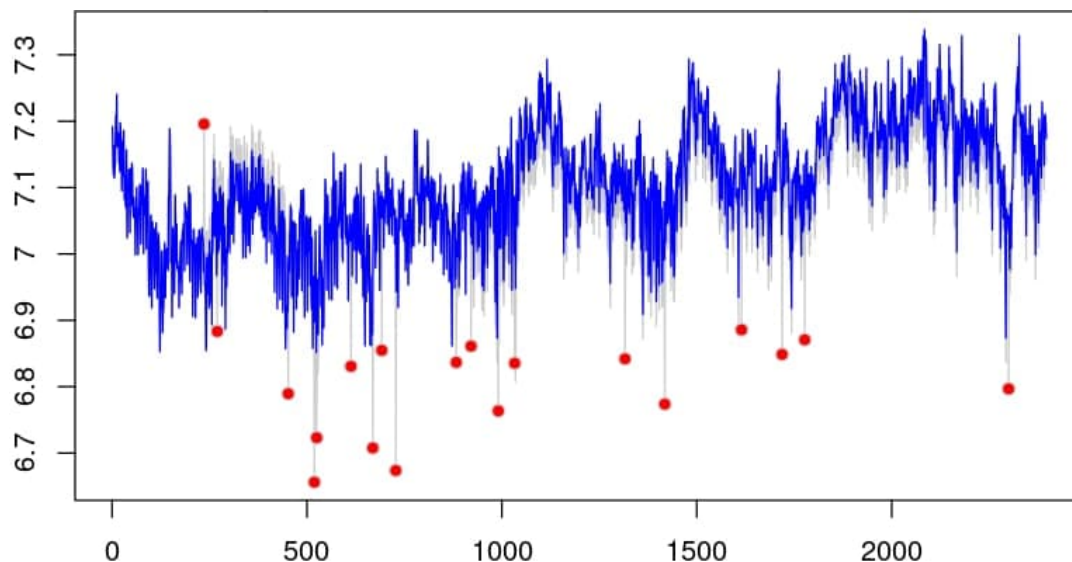
Parameter	n_{eff}	mean	sd	2.5%	50%	97.5%
ccdd	32	0.040	0.045	0.003	0.022	0.137
chdd	23	0.087	0.041	0.032	0.087	0.204
ctempp	33	-0.010	0.005	-0.020	-0.009	-0.004
ctempn	35	0.002	0.003	-0.005	0.002	0.009
ctemppl	26	0.008	0.004	0.003	0.007	0.016
ctempnl	33	-0.002	0.004	-0.011	-0.002	0.006
csat	43	0.013	0.010	-0.005	0.011	0.033
csun	36	-0.012	0.007	-0.028	-0.012	0.002
emon	35	0.010	0.009	-0.002	0.008	0.029
chol	135	0.018	0.013	-0.011	0.020	0.046
ctrend	24	0.006	0.005	0.001	0.005	0.018
cipp	33	2.037	1.105	0.523	1.924	3.963
cgas	74	0.001	0.001	0.000	0.001	0.002
ccoal	42	0.008	0.005	0.001	0.008	0.018
mu	29	6.529	0.466	5.791	6.424	7.638
kap	3	-2.911	0.812	-4.022	-3.023	-1.807
phi	2	0.114	0.087	-0.034	0.107	0.231
nu_u	25	0.352	0.162	0.099	0.326	0.795
nu_d	37	761.766	639.793	65.557	602.897	2452.5

Примечание – значимые на 5% уровне коэффициенты с априорными распределениями на всей оси $(-\infty, \infty)$ и со средним в 0 отмечены жирным шрифтом.

Источник: составлено авторами.

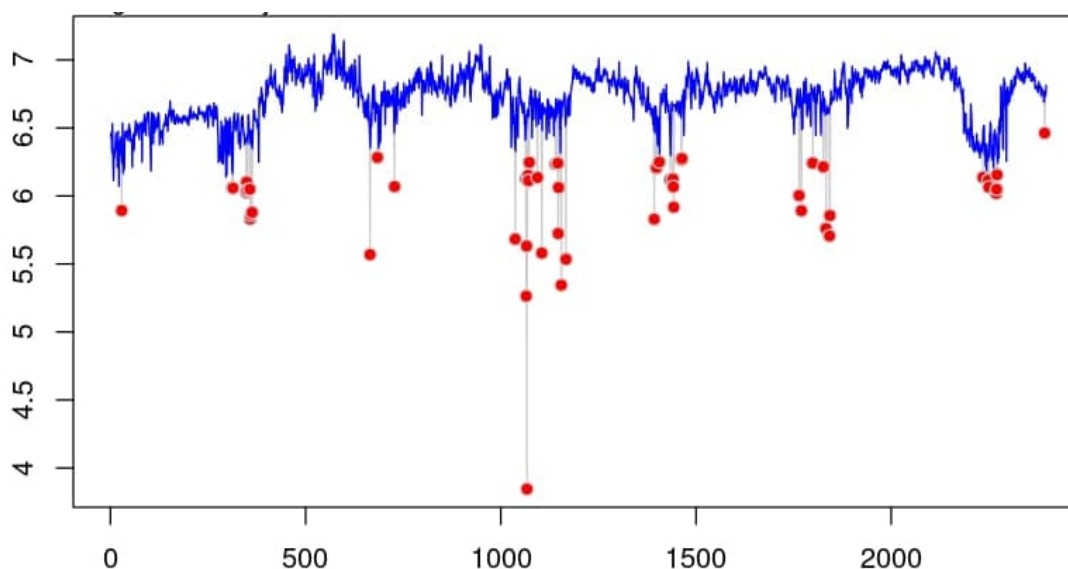
Однако, стоит отметить, что для обеих моделей коэффициенты при ВВП, коэффициент при тренде, а также коэффициенты при ценах на ресурсы являются статистически значимыми, поскольку учет скачков позволил уменьшить стандартные ошибки оценок коэффициентов. Причем знаки оценок при этих репрессорах совпадают с предположениями, сделанными по модели.

На рис. 10-11 представлены оцененные моменты скачков, то есть переменная q_t .



Примечание – Источник: составлено авторами.

Рис. 10. Скачки выявленные по модели 5 в 1 ЦЗ



Примечание – Источник: составлено авторами.

Рис. 11. Скачки выявленные по модели 5 во 2 ЦЗ

Можно отметить, что и в 1 и во 2 ЦЗ подавляющее большинство выявленных скачков – это скачки вниз (в первой модели был выявлен только 1 скачок вверх, во второй модели – скачков вверх выявлено не было). Наличие таких скачков объясняется недостатком спроса в эти дни.

Основным результатом данной модели является то, что добавление скачкообразной компоненты в модель значительно ухудшает показатели надежности байесовских оценок. При этом можно заметить, что учет скачков в модели необходим, поскольку точность оценок повысилась.

Отметим, что в модели, где учитываются скачки было обнаружено положительное влияние охлаждающего периода на цены. Причем, влияние этой переменной выше для 2 ценовой зоны. В частности, во время отопительного периода цены на электричество в среднем выше на 0.05% в первой ценовой зоне, и на 0.08% выше в второй. Тот факт, что оценки по модулю меньше 1 может объясняться тем, что часть среднегодовой динамики цен была учтена с помощью сезонных регрессоров, моделируемых как ряды Фурье.

Также, стоит отметить, что в обеих ценовых зонах коэффициент при тренде положительный и значимый, причем, в первой ценовой зоне он составил 0.004% в день, или 1.46% в год, а во второй ценовой зоне – 0.006% в день, или 2.19% в год. Полученные значения годовой инфляции цен на электричество значительно ниже показателей инфляции, однако надо учесть, что в модели также фигурирует переменная тренда ИПП, а также цены на ресурсы, которые также имеют положительные и значимые оценки при них, поэтому данная оценка инфляции цен на электричество отражает только рост цен, связанных с ненаблюдаемыми в модели факторами.

Так в обеих моделях наблюдается положительное и статистически значимое влияние трендовой компоненты ИПП. Причем в первой ценовой зоне в среднем при однопроцентном росте ИПП цены на электричество растут на 2.04 %, а во второй – на 0.9%. Также в обеих моделях наблюдается положительное и статистически значимое влияние цен на ресурсы. Причем в первой ценовой зоне в среднем при однопроцентном росте цены на газ цены на электричество растут на 0.001 %, а во второй – на 0.001%. В первой ценовой зоне в среднем при однопроцентном росте цены на уголь цены на электричество растут на 0.008 %, а во второй – на 0.002%. Как и предполагалось, во второй ценовой зоне цены на уголь имеют положительное влияние на цены на электроэнергию и оно выше, чем влияние цен на газ.

Вывод по оцениванию Модели 5 – необходимо либо предварительно скорректировать выбросы, либо добавить дамми переменную на наличие скачка, либо использовать модель с робастным к выбросам распределением.

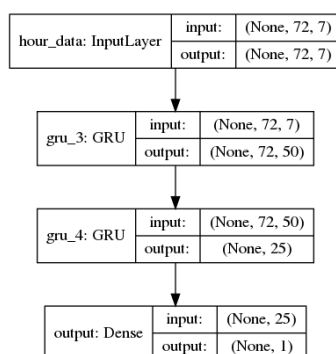
3.2. Прогнозирование почасовых объемов спроса на электроэнергию на рынке на сутки вперед с использованием разработанной модели

В качестве основной модели прогнозирования используется предложенная выше нейронная сеть, совмещающая сверточные и рекуррентные блоки (CNN-RNN), где размерность месячных данных приводится к редуцированной размерности часовых данных посредством встроенной полносвязной сети перед блоком конкатенации. Один обучающий набор такой сети представляют две двумерные матрицы. Первая матрица – это временное окно часовых данных, ширина которого равна количеству объясняемых переменных, имеющих часовое измерение, а длина – 72 точки, что соответствует трем дням наблюдений непосредственно до значения целевой переменной. Вторая матрица – это временное окно месячных данных, его ширина, соответственно, равна количеству объясняющих переменных с месячным измерением, длина – 3 значения. При этом функция разбиения данных на обучающие шаблоны устроена так, что три лаговых наблюдения месячных данных будут объяснять прогнозируемое целевое значения в любой день текущего месяца. К каждому сверточному слою применяется регуляризация «kernel_regularizer», которая позволяет накладывать штраф за высокие значения коэффициентов в ядре свертки, что снижает риск переобучения и вынуждает информацию внутри нейронной сети распространяться более равномерно. Кроме того, к выходным данным каждого сверточного и рекуррентного слоя применяется батч-нормализация. Этот блок нормализует выходные данные каждого слоя, что позволяет сгладить ландшафт функции потерь и ускорить процесс обучения¹, а также снизить риск искусственного завышения или занижения значимости определенного фактора (из-за возможной разницы в уровнях рядов). Во внутренних слоях используются функции активации «gelu». Последний, выходной слой представляет собой сумматор с линейной функцией активации. Изначально свободные члены инициализируются нулями, остальные параметры нейронной сети с использованием «HeNormal» - нормального распределения. В процессе обучения нейронной сети среднеквадратическая ошибка отклонения на обучающей выборке служит для расчета значения целевой функции; ошибка на валидации – для определения гиперпараметров. В том случае, если в течение трех эпох не происходит снижения валидационной ошибки, шаг градиентного спуска сокращается. Вся выборка может использоваться для обучения до 30 раз (количество эпох), но обучение останавливается, если на протяжении 5-ти эпох ошибка на валидационной выборке не уменьшается.

¹ Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How does batch normalization help optimization?, Advances in Neural Information Processing Systems

Для проверки оптимальности конструкции предложенной модели также используются две нейронных сети с базовой исходной архитектурой, но исключенными и измененными некоторыми блоками. Первая такая модель предполагает замену блока преобразования месячных данных с использованием инкорпорированной полносвязной нейронной сети на механизм «UpSampling1D», который также приводит к изменению размерности месячных данных до редуцированной размерности часовых данных, но делает это более линейно и задействует на порядок меньше параметров. Вторая модель предполагает исключение блока с месячными данными. Таким образом, в рамках упрощенной модели предполагается, что одно прогнозное значение – это функция от трехдневного временного окна почасовых данных, которые включают в себя лаговые значения целевой переменной и синтетические переменные, отражающие сезонную структуру данных.

В качестве бенчмарков используется рекуррентная нейронная сеть и сезонная ARIMA с экзогенными факторами. Архитектура рекуррентной сети представлена на рис. 12.



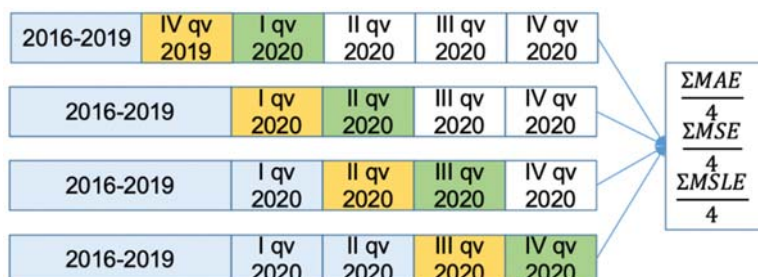
Примечание – Источник: составлено авторами.

Рис. 12. Архитектура рекуррентной нейронной сети, используемой в качестве альтернативной модели

Рекуррентная сеть имеет два скрытых слоя. На первом скрытом слое находится 50 рекуррентных ячеек GRU (Gated recurrent units), которые возвращают временные последовательности. На втором скрытом слое находится 25 рекуррентных ячеек, каждая из которых возвращает единственное значение. На последнем слое сумматор линейно взвешивает выходы предыдущего рекуррентного слоя, что и является прогнозным значением нейронной сети, которое используется для расчета ошибки в процессе обучения. Во внутренних слоях также используется функция активации «Relu», а выходы внутренних слоев нормализуются, как и в случае основной архитектуры.

Расчет ошибки основной и альтернативных моделей осуществляется с учетом валидации, которая, по сути, предполагает разделение выборки на обучающую, валидационную

и тестовую несколькими способами. Валидация выборки с временными рядами необходима, поскольку обобщающая способность нейросети, обученной с использованием одного, определенного валидационного множества, может быть ограниченной. Для того, чтоб проверить устойчивость модели мы используем схему обучения, приведенную на рисунке 13.



Примечание – Источник: составлено авторами.

Рис. 13. Валидация выборки в процессе обучения нейронной сети

Приведенная схема основана на стандартной практике валидации временных рядов со сдвигом временного окна. В соответствии со схемой, каждая модель обучается четыре раза, используя каждый раз сдвинутые на квартал вперед данные. На блоках с голубым фоном обозначены тренировочные множества модели. Эти данные используются непосредственно для обучения основных параметров нейронной сети. Желтым фоном обозначены валидационные множества. По этим данным определяется момент остановки процесса обучения и точка экстремума функции ошибки с предполагаемой максимальной обобщающей способностью. Зеленым фоном обозначены тестовые множества. Эти данные скрыты на всех этапах обучения и используются только для расчета ошибки. На белом фоне обозначены отложенные данные для следующих итераций расчетов.

После каждого этапа обучения рассчитывается три вида ошибки:

- MAE (Mean absolute error) – средняя абсолютная ошибка,
- MSE (Mean squared error) – среднеквадратичная ошибка,
- MSLE (Mean squared logarithmic error) – среднеквадратичная логарифмическая ошибка.

Последнюю можно интерпретировать, как меру соотношения между истинными и прогнозируемыми значениями. Это означает, что MSLE обрабатывает небольшие отклонения между небольшими истинным и прогнозным значением приблизительно также, как большие различия между большими истинными и прогнозируемыми значениями. Кроме того, такая ошибка предполагает больший штраф за недооценку, чем за завышение. В случае прогнозирования объемов электроэнергии данную ошибку имеет смысл использовать,

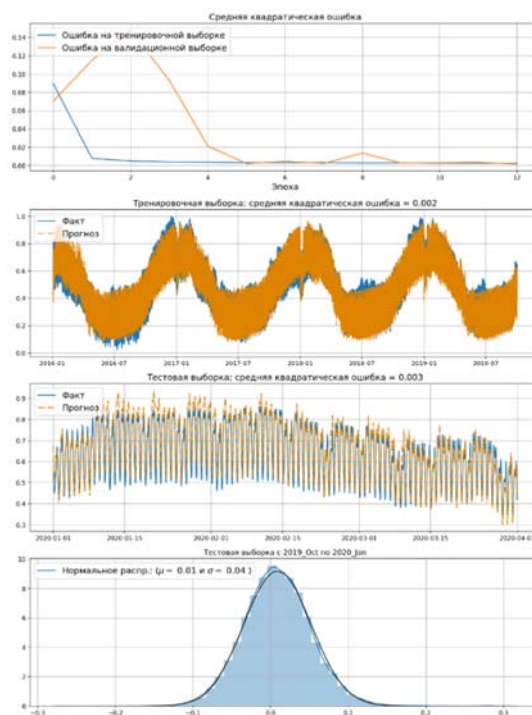
как метрику качества модели, поскольку разброс целевой переменной значителен (в рабочее время объемы потребляемой энергии значительно выше, чем в нерабочие и праздничные дни). К тому же, с точки зрения негативных последствий не оправдания прогнозов, социальные риски гораздо ниже от возможного перепроизводства электроэнергии, чем недопроизводства объема, на который предъявлен спрос.

Сравнение рассматриваемых моделей между собой производится по усредненным за 4 квартала ошибкам, а также с применением теста Diebold-Mariano на статистически значимое отличие ошибок. Кроме того, для каждого этапа обучения рассчитывается коэффициент детерминации, отражающий долю объясненной дисперсии истинной объясняемой переменной.

На рисунке 14 приведены результаты обучения основной нейросетевой модели CNN-RNN с совмещением данных разной частотности, где в качестве валидационного множества используется данные четвертого квартала 2019 года, а в качестве тестового множества – данные первого квартала 2020 года. Первый график рисунка иллюстрирует процесс обучения нейронной сети. В данном случае сеть обучилась за 12 эпох, поскольку в течении 5 последних не происходило уменьшения ошибки на валидации. Кривая ошибки на тестовом множестве монотонно убывает, однако кривая ошибки на валидационных начала снижаться только после первых двух эпох. Это может говорить о том, что первоначальная инициализация весов была не совсем удачна, но в процессе обучения модель все же достигла минимума, который обеспечил удовлетворительную обобщающую способность сети. Об этом свидетельствует незначительно отличающаяся среднеквадратическая ошибка на тестовых данных, которые были полностью скрыты от нейронной сети в процессе обучения, от ошибки на тренировочной выборке. Второй рисунок иллюстрирует качество внутривыборочного прогноза со среднеквадратической ошибкой 0.002. Третий график иллюстрирует качество вневыборочного прогноза. На тестовых данных значение ошибки составило 0.003. Последний график демонстрирует распределение остатков модели. Судя по графику, можно сказать, что остатки имеют нормальное распределение с математическим ожиданием 0.01 и среднеквадратическим отклонением 0.04. Важно отметить, что приведенные на графике расчеты произведены для масштабированных исходных данных.

Преимуществом сверточных слоев является возможность их простой интерпретации посредством изучения структуры весов фильтров. Некоторое понимание используемых закономерностей нейронной сетью для прогноза может обеспечить изучение самого первого сверточного слоя, который преобразует входные данные. Рассмотрим блок нейронной сети, отвечающий за обработку часовых данных. В данном случае первоначальная размерность входных данных – 72 наблюдения по оси времени и 7

факторов: лаговое значение объема производства; лаговое значение цены; дамми на государственные праздники; номер недели года; номер дня недели года; синус времени в формате «Unix time»², совершающий одно полное колебание в течении недели; синус времени, совершающий одно полное колебание в течении дня.



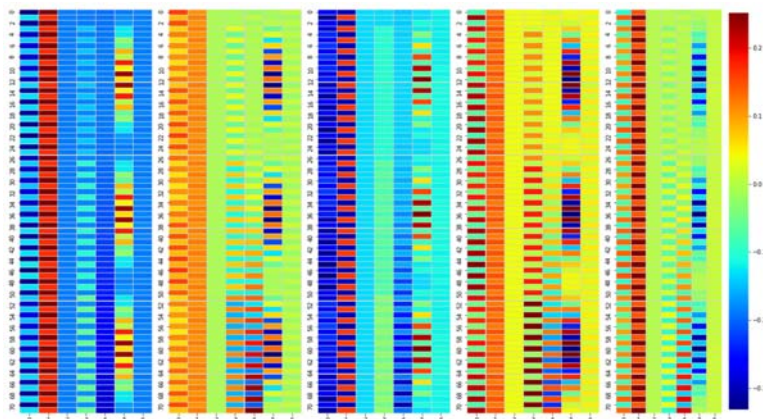
Примечание – Источник: расчеты авторов.

Рис. 14. Результат обучения CNN-RNN модели с тестовыми данными первого квартала 2020 года

Стоит заметить, что последние четыре фактора попарно несут в себе один и тот же смысл – учет внутринедельной и внутридневной сезонности. Только первые два имеют дискретный, ломаный график, а вторые два более гладкий, непрерывный. Однако, поскольку исключение одного из них приводило к увеличению тестовой ошибки, то в конечной модели было решено использовать факторы одновременно. Общее число фильтров на первом слое равно 15 (filters=15), размерность сверточного ядра – (2, 7), шаг ядра – 2 значения по оси времени (strides=2). Если привести матрицу входных данных к размерности (36, 2, 7), то есть, фактически разделить на 36 шаблонов и каждый скалярно умножить на матрицу основных весов сверточного ядра, то в результате получится некоторая карта важности исходных факторов. Описанная операция является первым действием так называемой операции свертки, после которой происходит суммирование значений каждой из 36 матриц с добавлением обучаемого параметра смещения (изначально они заданы нулями). Затем полученные значения используются нелинейной функцией

² Количество секунд, прошедших с полуночи 1 января 1970 года.

активации, в данном случае «Relu», в результате чего формируется матрица размерности (36, 15), каждый столбец которой – это результат поочередного применения 15 одномерных сверточных фильтров к исходным данным. Данная матрица и есть выход сверточного слоя и одновременно вход следующего слоя сети. На рис. 15 представлены тепловые карты важности исходных факторов для первых пяти сверточных фильтров, рассчитанные для первого шаблона тестовых данных.



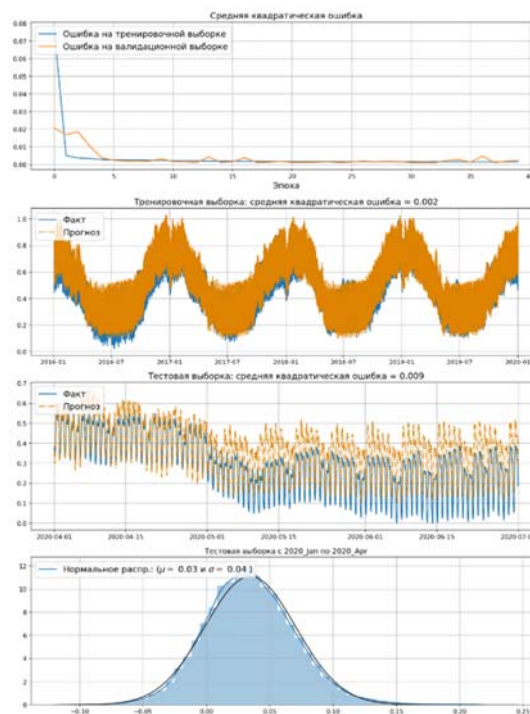
Примечание – Источник: расчеты авторов.

Рис. 15. Тепловые карты важности исходных факторов первых 5 сверточных слоев нейронной сети

По оси ординат тепловой карты располагаются 72 лага обучающих данных от целевой переменной. По оси абсцисс – семь используемых факторов, перечисленных выше. Поскольку далее все значения по оси абсцисс складываются, а результирующее значение – это смещенное на константу значение элемента вектора, подающегося на вход следующему слою, то можно сказать: чем выше значение на тепловой карте у признака, тем этот признак вносит больший вклад в построенное нейронной сетью прогнозное значение, соответствующее рассматриваемому шаблону. По приведенным выше тепловым картам можно заметить, что лаг цены (обозначен номером 1 на оси абсцисс) практически во всех фильтрах имеет весомый вклад в формирование прогноза, поскольку его активации достаточно высоки; синтетические переменные на основе синусоидальных функций, отражающие внутринедельные и внутридневные особенности колебания спроса на электроэнергию, также имеют значимый вклад в прогноз. При этом, поскольку приведенная тепловая карта иллюстрирует важность факторов для одного конкретного прогноза в некоторый период времени, то также можно заметить, на какие лаги из 72-часового окна нейронная сеть обращает внимание. Фактор с лаговыми значениями объемов производства электроэнергии используется 4 и 5 фильтрами, как достаточно информативный. При этом, можно заметить, что каждый второй лаг в рассматриваемом окне считается несущественным, поскольку значе-

ния его активаций уходят в отрицательную область. Это можно объяснить случайной инициализацией, в результате которой соответствующий вес был равен нулю и не изменился в процессе работы алгоритма обратного распространения ошибки или результатом используемой регуляризации, которая накладывает штраф за высокие значения параметров. В любом случае, использование всех лагов фактора считается сетью избыточным. В данном случае, возможно, из-за высокой автокорреляции первого порядка, в результате которой два соседних значения ряда несут приблизительно один объем полезной для прогнозирования информации.

На рис. 16 приведены результаты обучения основной нейросетевой модели, где в качестве валидационного множества используется данные первого квартала 2020 года, а в качестве тестового множества – данные второго квартала 2020 года. В данном случае процесс обучения занял максимально предусмотренные 40 эпох. Это связано с нестабильным поведением функции ошибки на валидации. После резкого снижения ошибки до нуля по прошествии четырех эпох, возможно, был обнаружен некоторый неустойчивый локальный минимум, из которого периодически стохастичность градиентного спуска позволяла оптимизатору выбраться, что отражает систематически повышающаяся ошибка, но более глубокий оптимум, скорее всего, так и не был найден. Однако в результате, ошибка на обучающей выборке достигла такого же уровня, как и в предыдущем случае – 0.002. Но на тестовой выборке значительно выросла – до 0.009.



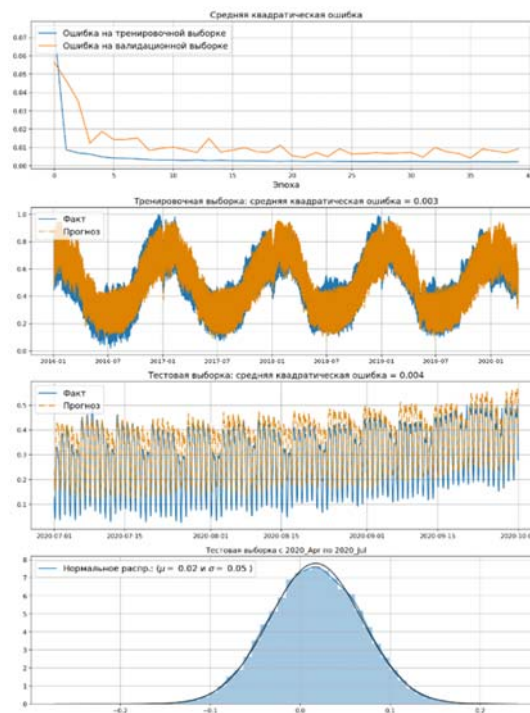
Примечание – Источник: расчеты авторов.

Рис. 16. Результат обучения CNN-RNN модели с тестовыми данными второго квартала

2020 года

Рост ошибки на тестовой выборке объясняется наличием структурного сдвига, отчетливо наблюдающегося на 3 графике в уровне объемов произведенной электроэнергии в мае 2020 года. Структурный сдвиг произошел по причине начала интенсивного распространения коронавирусной инфекции, повлекшего за собой введение ряда ограничительных эпидемиологических мер, в том числе введение карантина и закрытие границ. Последнее привело к разрыву производственных цепочек и снижению объемов производства предприятий, зависящих от импорта или напротив, экспортирующих активно свою продукцию. В результате спрос на электроэнергию снизился нехарактерным для данного сезона образом. Нейронная сеть, как и любая математическая модель, основанная на выявленных закономерностях в прошлом, подобный исход не могла учесть. Отсюда и систематическое завышение прогноза после мая 2020 года.

На рис. 17 приведены результаты обучения основной нейросетевой модели, где в качестве валидационного множества используется данные второго квартала 2020 года, а в качестве тестового множества – данные третьего квартала 2020 года.



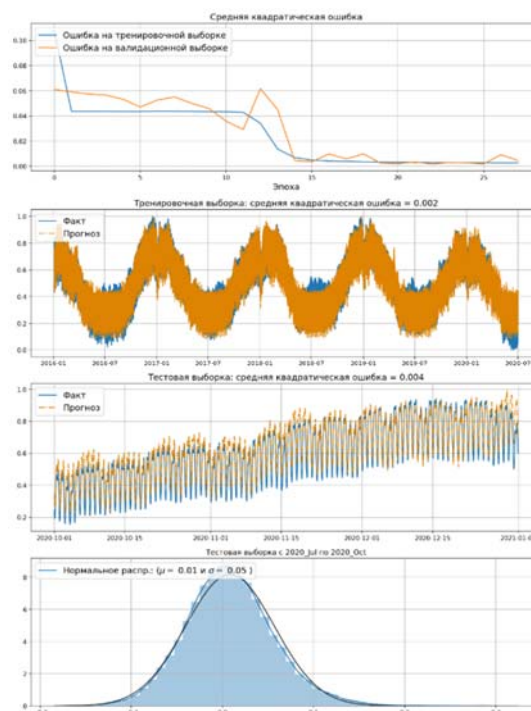
Примечание – Источник: расчеты авторов.

Рис. 17. Результат обучения CNN-RNN модели с тестовыми данными третьего квартала 2020 года

В данном случае завышенный прогноз на тестовой выборке также является результатом неучтенных факторов в виде начала коронавирусной эпидемии и ее последствий.

Стоит отметить, что в случае использования дамми переменной для идентификации структурного сдвига – самого простого способа учесть влияние пандемии, результат обучения и прогнозирования в данном дизайне эксперимента не мог быть изменен, поскольку дамми со значением 1, обозначающим наличие экзогенного шока, попадала бы только на валидационную выборку, которая прямо не участвует в процессе обучения. Однако в случае реального применения разработанной модели валидационную и тестовую выборку было бы уместно уменьшить, тем более учитывая, что целевая переменная имеет часовое измерение и наличие 2 160 точек (среднее число часов в квартале) явно излишне для тестирования с учетом ориентированности модели на краткосрочное прогнозирование. В таком случае, часть данных с периодом, включающим наличие эпидемии, могло оказаться в обучающей выборке, и добавленная дамми могла бы учесть влияние данного шока. Однако для более явного подсвечивания возможных проблем с использованием нейронных сетей в качестве прогнозных моделей, мы сохраняем исходно выбранные гиперпараметры.

На рис. 18 приведены результаты обучения основной нейросетевой модели, где в качестве валидационного множества используются данные третьего квартала 2020 года, а в качестве тестового множества – данные четвертого квартала 2020 года.



Примечание – Источник: расчеты авторов.

Рис. 18. Результат обучения CNN-RNN модели с тестовыми данными четвертого квартала 2020 года

В данном случае по-прежнему не учитывалось влияние пандемии коронавируса в явном виде для сохранения принципа при прочих равных. Однако стоит отметить, что ла-

говые значения целевой переменной, цен на электроэнергию, макроэкономических месячных данных уже включали в себя результаты влияния пандемии. При этом к декабрю 2020 года экономика, возможно, адаптировалась к новым условиям. Вневыборочный прогноз в этом периоде уже не отличается значительным завышением.

В таблице 6 приведены ошибки на тестовой выборке, а также коэффициенты детерминации модели на каждом оцениваемом периоде.

Лучше всего модель объясняет изменение объемов производства в первом и четвертом кварталах 2020 года (процент объясненной дисперсии целевой переменной соответственно составляет 90.2% и 92%). Хуже всего модель объясняет динамику объемов производства электроэнергии во втором квартале 2020 года, что связано с началом эпидемии коронавируса – 68% дисперсии. В третьем квартале экономика еще не полностью приспособилась к возникшим ограничениям в условиях пандемии, в результате чего модель, обученная на данных без воздействия данного шока, показала точность несколько выше, чем во втором квартале – 78% объясненной дисперсии. Однако в четвертом квартале, возможно, произошла адаптация экономики к новым условиям, в результате чего модель, предполагающая неизменность основных экономических законов, по которым взаимодействуют между собой объясняющие и целевая переменная, вновь продемонстрировала относительно высокую точность вневыборочного прогноза.

Таблица 6

Метрики эффективности основной нейросетевой модели на тестовой выборке

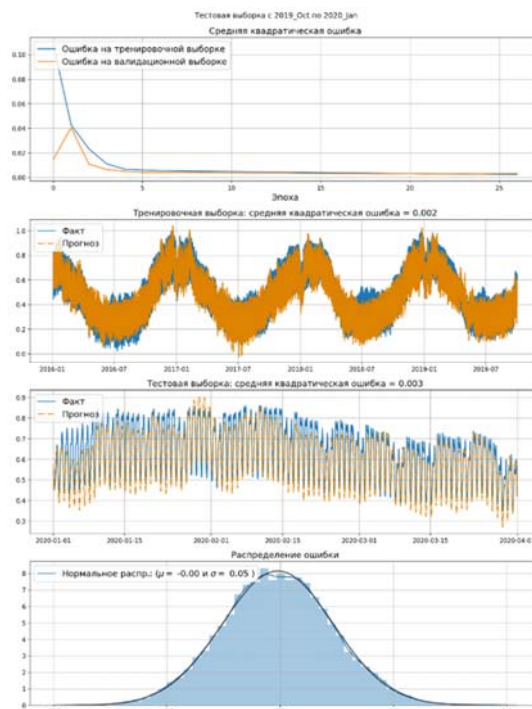
Модель	Квартал	MAE	MSE	MSLE	R2
CNN-RNN	1	0,031	0,003	0,001	0,902
	2	0,067	0,009	0,004	0,678
	3	0,052	0,006	0,002	0,783
	4	0,045	0,003	0,001	0,92
	Среднее	0,049	0,005	0,002	0,821

Примечание – Источник: расчеты авторов.

Далее рассмотрим процесс обучения и результаты прогнозирования вне выборки простой рекуррентной нейросетевой модели, обученной поочередно на аналогичных четырех подвыборках. На рис. 19 приведены результаты обучения, где в качестве валидационного множества используются данные четвертого квартала 2019 года, а в качестве тестового множества – данные первого квартала 2020 года.

Стоит отметить, что рекуррентная модель не предполагает совмещения данных разной размерности, но использует интерполированные линейно месячные данные среди объясняющих переменных для прогноза. В данном случае более простая модель демонстрирует сопоставимые по качеству прогнозы с моделью CNN-RNN: среднеквадратические ошибки

на обучающей и тестовой выборках вплоть до третьего знака не отличаются между собой, хотя визуально прогноз более простой модели вне выборки, кажется, имеет систематическое занижение. Для того, чтоб ответить более точно на вопрос – какая модель более эффективна, в конце главы воспользуемся статистическим тестом на значимость отличия ошибок между основной и альтернативной моделями.



Примечание – Источник: расчеты авторов.

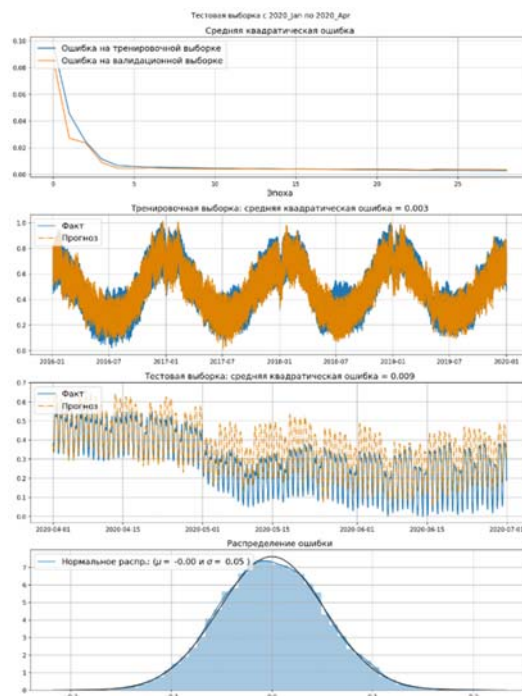
Рис. 19. Результат обучения RNN модели с тестовыми данными первого квартала 2020 года

На рис. 20 приведены результаты обучения рекуррентной нейросети, где в качестве валидационного множества используются данные первого квартала 2020 года, а в качестве тестового множества – данные второго квартала 2020 года.

В данном случае более простая нейросеть демонстрирует несколько более высокую ошибку на обучении, но при этом сопоставимую ошибку на тесте. Как и в случае основной модели, рекуррентная сеть плохо учитывает структурный сдвиг в данных, вызванный наступлением коронавирусной эпидемии. На рисунке 21 приведены результаты обучения рекуррентной нейронной сети, где в качестве валидационного множества используются данные второго квартала 2020 года, а в качестве тестового множества – данные третьего квартала 2020 года.

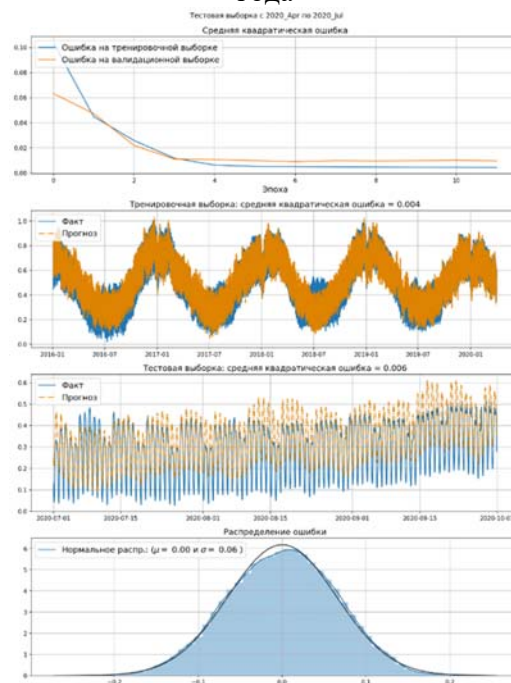
В данном случае более простая нейросетевая модель демонстрирует результаты хуже на обучающей и тестовой выборке, чем основная модель, согласно значениями среднеквадратических ошибок.

На рис. 22 приведены результаты обучения рекуррентной нейронной сети, где в качестве валидационного множества используются данные третьего квартала 2020 года, а в качестве тестового множества – данные четвертого квартала 2020 года.



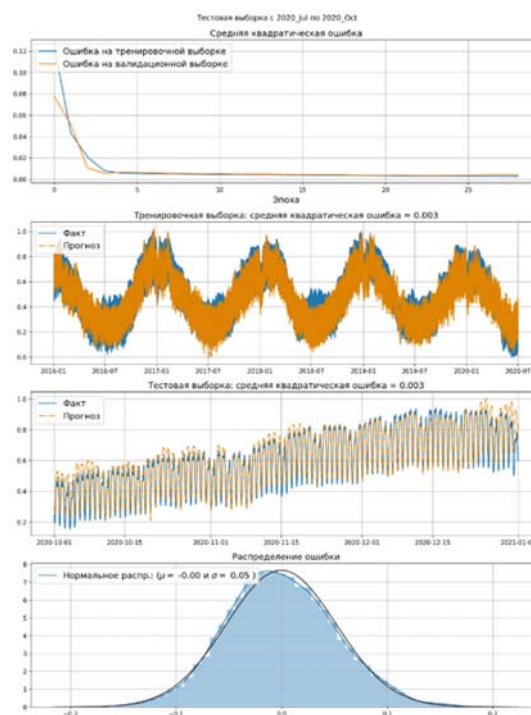
Примечание – Источник: расчеты авторов.

Рис. 20. Результат обучения RNN модели с тестовыми данными второго квартала 2020 года



Примечание – Источник: расчеты авторов.

Рис. 21. Результат обучения RNN модели с тестовыми данными третьего квартала 2020 года



Примечание – Источник: расчеты авторов.

Рис. 22. Результат обучения RNN модели с тестовыми данными четвертого квартала 2020 года

По результатам обучения в данном случае ошибки на обучающей и тестовой выборках совпадают до третьего знака. В таблице 7 приведены три вида ошибок для рекуррентной модели, обученной на данных со сдвигом временного окна в один квартал.

Таблица 7

Метрики эффективности рекуррентной нейросетевой модели на тестовой выборке

Модель	Квартал	MAE	MSE	MSLE	R2
RNN	1	0,031	0,003	0,001	0,906
	2	0,063	0,009	0,003	0,725
	3	0,073	0,004	0,004	0,609
	4	0,038	0,004	0,001	0,932
	Среднее	0,051	0,005	0,002	0,793

Примечание – Источник: расчеты авторов.

Согласно результатам, как и в случае основной модели, в первый и четвертый кварталы модель продемонстрировала наиболее высокие показатели точности прогноза и соответственно низкие ошибки на тестовой выборке.

Для того, чтоб проверить оптимальность предложенной архитектуры CNN-RNN были проведены аналогичные, представленным выше, расчеты на подвыборках с упрощенными по строению нейронными сетями, в основе которых находится базовая модель. В первом случае конструктивные измерения были внесены в блок обработки месячных данных, которые преобразует входящую матрицу с низкой размерностью в матрицу с более высокой

размерностью. В основной модели этот блок был реализован посредством полносвязной нейронной сети, которая имела один слой и число нейронов в этом слое соответствовало размерности, к которой необходимо было привести исходные месячные данные. Измененная модель предполагает использование механизма «UpSampling1D» для цели увеличения размерности матрицы месячных данных. В отличие от инкорпорированной полносвязной нейронной сети, данный механизм имеет более линейную логику и задействует на порядок меньше параметров в процессе преобразования данных. В таблице 8 приведены результаты обучения такой модели.

Таблица 8

Метрики эффективности основной нейросетевой модели с блоком «UpSampling1D»

Модель	Квартал	MAE	MSE	MSLE	R2
CNN-RNN UpSampling1D	1	0,03	0,003	0,001	0,915
	2	0,064	0,008	0,003	0,699
	3	0,065	0,005	0,003	0,701
	4	0,035	0,003	0,001	0,949
	Среднее	0,049	0,005	0,002	0,816

Примечание – Источник: расчеты авторов.

Модель с иной конструкцией блока обработки месячных данных в целом показала схожий результат с основной моделью, согласно средним ошибкам на тестовой выборке, но при этом средний коэффициент детерминации у первоначальной модели несколько выше: 0.821 против 0.816.

Во втором случае основная модель была значительно упрощена. Из первоначальной архитектуры был удален блок, связанный с использованием месячных данных. Целью такого эксперимента является проверка возможностей низкоразмерных данных нести дополнительную, полезную информацию. В том случае, если результирующие метрики основной модели хотя бы в случае некоторых тестовых множеств превосходят метрики редуцированной модели, то можно сделать вывод о необходимости использования рассматриваемых низкоразмерных объясняющих переменных для полноты модели. В таблице 9 приведены результаты обучения модели без использования месячных данных с сохранением прочих особенностей строения базовой архитектуры предлагаемой CNN-RNN модели.

Согласно результатам, упрощенная модель без использования месячных данных, на трех кварталах из четырех демонстрирует более высокую ошибку и одновременно относительно низкий R^2 по сравнению с полной моделью. Это говорит о том, что использование месячных данных позволяет нейронной сети задействовать дополнительную информацию для результирующего вывода, которую не содержат высокочастотные объясняемые переменные. Возможно, это связано с тем, что низкочастотные данные, в данном случае – ин-

дексы промышленного производства наиболее энергоемких отраслей, могут нести информацию о конъюнктуре рынка, которая влияет на прогнозируемое значение уровня тренда³. В то время, как сезонная динамика целевой переменной в большей степени объясняется синтетическими переменными для учета сезонности и лаговыми значениями цены и объемов производства электроэнергии.

Таблица 9

Метрики эффективности основной нейросетевой модели без использования низкоразмерных данных

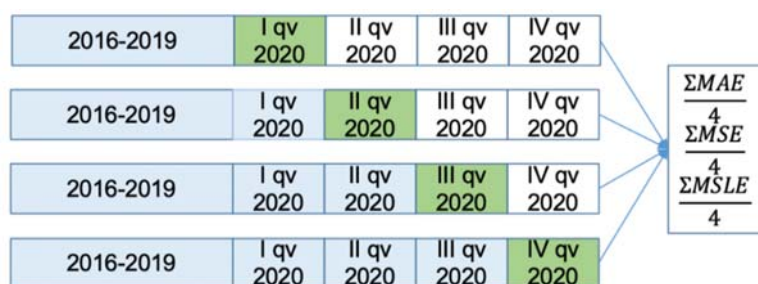
Модель	Квартал	MAE	MSE	MSLE	R2
CNN-RNN без месячных данных	1	0,042	0,003	0,001	0,896
	2	0,067	0,009	0,005	0,607
	3	0,054	0,006	0,003	0,789
	4	0,051	0,004	0,002	0,889
	Среднее	0,054	0,006	0,003	0,795

Примечание – Источник: расчеты авторов.

В качестве простого линейного бенчмарка выступает сезонная ARIMA с экзогенными переменными, среди которых находятся, в том числе, квартальные данные, интерполированные линейно до большей размерности.

Для выбора оптимальных параметров модели используется полный перебор 81 комбинации (p, d, q) и сезонных параметров, где p – число AR лагов, d – степень дифференцирования ряда, q – число используемых MA лагов. При этом сезонный параметр имеет фиксированное значение и равен 7-ми. Оптимальной моделью признается модель с наименьшим значением информационного критерия Акаике. Выбор оптимальной модели с сохранением ее параметров производится четыре раза со сдвигом обучающего множества на один квартал вперед по аналогии с обучением нейросетевых моделей. При этом валидационная выборка в случае ARIMA относится к обучающей. Это связано с тем, что в процессе обучения нейронной сети валидационная выборка также принимает участие, хотя и неявным образом: по поведению ошибки на валидационном множестве принимается решение об остановке процесса обучения, то есть происходит непосредственно определение минимума функции ошибки, который предположительно обладает максимальной обобщающей способностью. Таким образом, для того, чтоб сравниваемые модели были построены на основе одинаковых информационных множеств, а также для наличия возможности сравнивать рассчитанные ошибки на аналогичных тестовых данных, валидация выборки в случае ARIMA модели имеет структуру, представленную на рис. 23.

³ Под трендом в данном случае понимается не тренд в строгом эконометрическом смысле объемов спроса, а компонента целевой переменной, представляющая собой очищенный исходный ряд от сезонности и случайного шума.



Примечание – Источник: расчеты авторов.

Рис. 23. Принцип валидации данных для авторегрессионной модели.

В результате, с использованием отобранных четырех оптимальных моделей на каждой обучающей выборке, согласно критерию Акаике, строятся вневыборочные прогнозы для каждого квартала 2020 года и рассчитывается матрица ошибок с коэффициентами детерминации. Прогноз строится итеративно. Результаты расчетов приведены в таблице 10.

Таблица 10

Метрики эффективности сезонной ARIMA модели

Модель	Квартал	MAE	MSE	MSLE	R2
SARIMA	1	0,068	0,023	0,021	0,695
	2	0,101	0,029	0,025	0,524
	3	0,084	0,026	0,023	0,645
	4	0,06	0,023	0,021	0,625
	Среднее	0,078	0,025	0,023	0,622

Примечание – Источник: расчеты авторов.

Ошибки линейной модели в среднем выше, чем у нелинейных, а процент объясненной дисперсии ниже в сопоставимых кварталах. При этом качество прогноза модели также, как и в других достаточно заметно снижается во втором квартале. Однако прогноз в третьем квартале превосходит качественно превосходит прогноз в 4 квартале, чего не наблюдалось в нелинейных моделях. Причиной этому скорее всего является то, что в нелинейных моделях данные второго квартала использовались в качестве валидационной выборки и не участвовали напрямую в процессе обучения, а линейной модели уже были доступны данные начала первой волны пандемии в обучающей выборке и параметры SARIMA были подобраны в соответствии с минимизацией информационного критерия на выборке с учетом структурных изменений. Для более наглядного сравнения моделей между собой, усредненные ошибки и коэффициенты детерминации за год – среднее по четырем кварталам, собраны в таблице 11.

Наименьшую среднеквадратическую ошибку прогноза ($MSE = 0.049$) и наиболее высокий процент объясненной дисперсии имеют CNN-RNN модели с двумя различными ме-

ханизмами обработки низкочастотных данных. Незначительно уступает в точности прогнозирования двухслойная рекуррентная нейронная сеть ($MSE = 0.051$). Несколько в большей степени уступает в точности прогнозирования нейронная сеть CNN-RNN без месячных данных ($MSE = 0.054$). Наибольшую среднюю ошибку прогноза (0.078) и одновременно низкий процент объясненной дисперсии (62%) имеет сезонная ARIMA.

Таблица 11

Сравнение усредненных по кварталам метрик рассматриваемых моделей

Модель	MAE	MSE	MSLE	R2
CNN-RNN	0,049	0,005	0,002	0,821
CNN-RNN UpSampling1D	0,049	0,005	0,002	0,816
CNN-RNN без месячных данных	0,054	0,006	0,003	0,795
LSTM	0,051	0,005	0,002	0,793
SARIMA	0,078	0,025	0,023	0,622

Примечание – Источник: расчеты авторов.

Поскольку усредненные ошибки различных моделей неотличимы на уровне десятичной дроби, то для сравнения моделей также применяется тест Diebold-Mariano (MD), нулевая гипотеза которого состоит в отсутствии различий между прогнозными свойствами двух моделей ($E[MSE_1^2 - MSE_2^2] = 0$). Особенностью MD теста является то, что функции потерь сравниваемых моделей не обязаны быть квадратическими или симметричными, кроме того, ошибки могут не подчиняться нормальному распределению, а также могут иметь ненулевое среднее и быть коррелированными серийно или одновременно. В том случае, если гипотеза о равенстве квадратов ошибок отвергается, отрицательная статистика теста показывает, что ошибка модели 2 больше ошибки модели 1, следовательно, модель 1 предпочтительнее модели 2. В таблице 12 приведены результаты теста, проведенного попарно с рассматриваемыми моделями.

В качестве тестовой выборки для расчета ошибок использовался первый квартал 2020 года. Статистика теста в каждом случае рассчитывалась с использованием среднеквадратической ошибки. В случае CNN-RNN модели с различными блоками обработки месячных данных разница двух ошибок статистически неотличима от нуля. Ошибки рекуррентной нейронной сети и CNN-RNN модели без блока месячных данных на 10% уровне значимости отличаются в большую сторону от ошибок базовой модели. В случае линейной модели гипотеза о равенстве двух ошибок отвергается, при этом ошибка линейной модели больше, чем ошибка нейросетевой. Таким образом, можно сделать вывод, что предложенная нейросетевая модель с совмещением данных разной частотности превосходит по качеству

прогноза линейный бенчмарк SARIMA и находится приблизительно на одном с рекуррентной нейронной сетью. При этом использование моделью низкоразмерных месячных данных в числе объясняемых переменных позволяет несколько повысить качество прогноза.

По результатам проведенных расчетов точность прогноза предложенной модели в нынешней конфигурации, измеряемая как средняя ошибка на тестовых выборках валидационных множеств, превосходит сезонную ARIMA и находится приблизительно на одном уровне с LSTM моделью. Кроме того, вариация архитектуры модели с отсутствующим блоком месячных макроэкономических факторов в числе объясняющих переменных продемонстрировала более низкие показатели точности, что подтверждает выводы из эконометрической модели стохастической волатильности о важности данных факторов в формировании динамики целевой переменной. При этом используемый сверточный механизм позволил использовать данные для прогноза без интерполяции, которая, возможно, вносит в данные дополнительный, неинтерпретируемый шум: в альтернативной модели с линейной интерполяцией месячных данных до размерности часовых ошибка на тесте отличалась в большую сторону.

Таблица 12

Сравнение рассматриваемых моделей при помощи теста

	CNN-RNN vs SARIMA	CNN-RNN vs LSTM	CNN-RNN vs CNN-RNN без месячных данных	CNN-RNN с блоком «UpSampling1D»
DM статистика	DM<0	DM<0	DM<0	DM<0
P-значение	0.01	0.07	0.09	0.38

Примечание – Источник: расчеты авторов.

Валидационный и тестовый периоды преднамеренно были выбраны продолжительностью в один квартал, что достаточно много с частотой измерения целевой переменной в один час. Однако это позволило подсветить возможные проблемы применения модели на практике в условиях отсутствия своевременной валидации. К примеру, наступление пандемии коронавируса значительно ухудшило точность получаемых прогнозов на период до двух кварталов, после чего экономика адаптировалась и прежняя модель, описывающая данные с удовлетворительной точностью, снова стала демонстрировать ошибку на прежнем, низком уровне. Это показало, что нейросетевая модель, как и любая математическая прогнозная модель, построенная на основе взаимозависимостей объясняемых факторов в прошлом, не может учитывать экзогенные шоки, до тех пор, пока они не отразятся в одной из независимых переменных и модель не будет переобучена с учетом данных, на которых произошел структурный сдвиг.

4. Интерпретация полученных результатов

Одним из приоритетных направлений деятельности государственных органов в настоящий период является достижение национальных целей развития, определенных в Указе Президента Российской Федерации от 07.05.2018 г. № 204 «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года», в частности, предполагается добиться существенного экономического роста. Учитывая существенную долю энергетического сектора в структуре ВВП России, сложившиеся тенденции на энергетическом рынке (рост доли зеленой энергетики в Европе, рост цен на нефть и природный газ, реализация государственной программы модернизации тепловых электростанций ДПМ-штрих, кризис, вызванный пандемией коронавирусной инфекции и другие) будут оказывать влияние на достижение национальных целей. Для учета такого числа факторов необходимо уметь строить достаточно сложные детализированные модели, учитывающие пространственные особенности рассматриваемой отрасли, разночастотность собираемой статистики, а также множество мелких факторов, которые могут быть ненаблюдаемы с точки зрения макростатистики (условная «стохастическая» компонента). Поэтому в рамках данной работы были построены модели, позволяющие учесть указанную выше специфику на рынке электроэнергетики.

Электроэнергия является товаром, который потребляется во всех областях реального сектора экономики. Соответственно динамика цен на электроэнергию напрямую влияет на издержки компаний и конкурентоспособность отечественной продукции, динамику цен, благосостояние потребителей и рост экономики.

Учитывая нелинейную природу динамики цен на электроэнергию⁴, а также наличие ненаблюдаемых факторов, оказывающих на нее влияние, моделирование и прогнозирование этих цен является сложной эмпирической задачей, одним из способов решения которой является применение моделей глубокого обучения, отбор факторов для которых осуществляется с использованием формальной эконометрической модели трендсезонного разложения, получаемого с использованием ценовых факторов спроса и предложения на рынке электроэнергии, а также с помощью добавления диффузионной компоненты, в которой учтены все ненаблюдаемые факторы, влияющие на цены.

⁴ Характерной особенностью, которую трудно учесть в традиционных методах моделирования, рынка электроэнергетики является то, что цены на электричество отличаются значительной волатильностью, обладают двухуровневой структурой сезонности (не считая дневной), отличаются наличием частых выбросов — скачков.

Существует множество различных подходов к моделированию цен на электричество. Традиционные методы временных рядов не могут точно моделировать нелинейную структуру цены на электроэнергию, также без учета скачкообразной компоненты доверительные интервалы могут быть заужены. В результате при анализе временных рядов цен на электроэнергию была выбрана гибкая модель, позволяющая учесть статистические особенности этих временных рядов: моделирование цены на оптовом рынке электроэнергии осуществлялось как диффузионно-скачкообразный процесс с учётом фундаментальных факторов со стороны спроса и предложения (температуры, уровня деловой активности и цен на используемое сырье при выработке электроэнергии).

На рынке электрической энергии большая часть объемов реализуется по свободным ценам, сформированным на рынке на сутки вперед (далее – РСВ). Формирование цен на РСВ осуществляется на основании заявок участников рынка по принципу маржинального ценообразования. Для каждого узла расчетной модели под воздействием спроса и предложения определяется цена на электрическую энергию, которая распространяется на всех участников. Ценовые заявки формируются поставщиками исходя из топливных затрат, так на динамику цен на рынке РСВ в значительной степени влияет изменение цен на топливо. В рамках моделирования рынка электроэнергии России была разработана модель ценообразования на электроэнергию на рынке на сутки вперед.

Результатом моделирования является тренд-сезонное разложение цены на электричество с интерпретируемыми оценками составляющих компонент. Помимо этого с помощью построенной модели можно определить причины различия в динамике цен в Европейской и Сибирской ценовых зонах (далее, как первая и вторая ценовые зоны), оценить темпы роста цен на электроэнергию, что немаловажно, поскольку, во-первых, расходы на электроэнергию несут подавляющее большинство производителей, во-вторых, сбытовые компании, участвующие на РСВ, в результате конкурентного отбора заявок, приобретают электричество, которое впоследствии будет отпускаться населению. Также оценка случайной компоненты внутри модели позволяет определить влияние структуры генерирующих мощностей на волатильность цен и оценить воздействие резких изменений экономических факторов на цены на электричество, а скачкообразная компонента, позволяет учесть неоднородность волатильности ошибок. Данная модель также может быть применена на равновесных поузловых ценах, что позволит использовать полученные оценки трендовой и сезонной составляющих, для выявления особенностей ценообразования в разрезе конкретных регионов России и разработки мер государственной политики в сфере регулирования рынка электроэнергии.

В результате оценки модели стохастической волатильности было получено, что дамми-переменные дней недели являются значимыми регрессорами. В первой ценовой зоне, как и предполагалось, в воскресные и праздничные дни цены на электричество ниже, чем в остальные дни. Во второй ценовой зоне, как и предполагалось, в воскресные дни цены на электричество ниже, чем в остальные дни. Влияния праздничных дней во второй ценовой зоне не выявлено, что не противоречит выдвинутым предположениям модели.

Такие результаты показывают различия в недельной динамике в первой и второй ценовой зоне. Основным отличием в недельной сезонности цен является в среднем более высокие цены в субботу во второй ценовой зоне при прочих равных по сравнению с более низкими ценами в субботу в первой ценовой зоне при прочих равных. Данный результат говорит о росте спроса на электричество в течение недели, причем во второй ценовой зоне этот рост наблюдается со вторника по субботу. Частично это может быть объяснено тем, что в первой ценовой зоне доля предприятий легкой промышленности выше, чем во второй, а во второй ценовой зоне выше доля предприятий непрерывного технологического цикла, работающих в таких отраслях, как металлургия, нефтегазовая отрасль. Предприятиям, занимающимся тяжелой промышленностью зачастую невыгодно останавливать производство, поэтому высокоэнергозатратные производства функционируют даже в выходные дни.

Данный результат можно, например, использовать при планировании графика ремонтных и профилактических работ на генерирующих объектах в каждой из ценовых зон.

Влияние дамми-переменной отопительного и охлаждающего периода, вероятно, было нивелировано включением в модель сезонности с помощью рядов Фурье. Но в результате оценки было обнаружено влияние отклонения температуры от плавающего среднего, причем в первой ценовой зоне сегодняшние отклонения от ожидаемой температуры, как вверх, так и вниз положительно влияют на цену, вчерашние отклонения от ожидаемой температуры, как вверх, так и вниз отрицательно влияют на цену, во второй ценовой зоне обнаружено только отрицательное влияние сегодняшнего отклонения температуры вверх, отрицательное влияние вчерашнего отклонения температуры вниз и положительное влияние вчерашнего отклонения температуры вверх. Данный результат говорит о необходимости планирования объемов потребления с учетом дневной температуры, точные прогнозы температуры позволят снизить волатильность цен на электричество.

В большинстве работ, рассмотренных в обзоре литературы, говорится именно о влиянии годовой или средней температуры на цены, связанной с охлаждательными и отопительными периодами года, но не учитывается резкие дневные колебания температуры, в чем и заключается новизна модели, построенной в данной работе.

Также, в результате оценивания модели было получено положительное влияние индекса промышленного производства и цен на ресурсы, выступающих в качестве топлива при генерации электроэнергии (природный газ, уголь). Причем во второй ценовой зоне влияние цен на уголь выше, чем влияние цен на газ. Таким образом, при прогнозировании инфляции цен на электроэнергию необходимо не только учитывать эти факторы, но и прогнозировать инфляцию отдельно в каждой из ценовых зон.

Выделенные компоненты разложения цены на электроэнергию в первой и второй ценовых зонах далее использовались, как факторы для прогнозирования равновесной цены на РСВ с помощью нейросетевой модели, совмещающей данные разной частоты.

По результатам проведенных расчетов точность прогноза предложенной конфигурации нейросети, измеряемая как средняя ошибка на тестовых выборках валидационных множеств, превосходит сезонную ARIMA и находится приблизительно на одном уровне с LSTM моделью. Кроме того, вариация архитектуры модели с отсутствующим блоком месячных макроэкономических факторов в числе объясняющих переменных продемонстрировала более низкие показатели точности, что подтверждает выводы из эконометрической модели стохастической волатильности о важности данных факторов в формировании динамики целевой переменной. При этом используемый сверточный механизм позволил использовать данные для прогноза без интерполяции, которая, возможно, вносит в данные дополнительный, неинтерпретируемый шум: в альтернативной модели с линейной интерполяцией месячных данных до размерности часовых ошибка на тесте отличалась в большую сторону.

Валидационный и тестовый периоды преднамеренно были выбраны продолжительностью в один квартал, что достаточно много с частотой измерения целевой переменной в один час. Однако это позволило выявить возможные проблемы применения модели на практике в условиях отсутствия своевременной валидации. К примеру, наступление пандемии коронавируса значительно ухудшило точность получаемых прогнозов на период до двух кварталов, после чего экономика адаптировалась и прежняя модель, описывающая данные с удовлетворительной точностью, снова стала демонстрировать ошибку на прежнем, низком уровне. Это показало, что нейросетевая модель, как и любая математическая прогнозная модель, построенная на основе взаимозависимостей объясняемых факторов в прошлом, не может учитывать экзогенные шоки, до тех пор, пока они не отразятся в одной из независимых переменных и модель не будет переобучена с учетом данных, на которых произошел структурный сдвиг.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Zdrozny P., "Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies," *Econometric Theory*, 1988. pp. 108-124.
2. Zdrozny P., "Estimating a multivariate ARMA model with mixed-frequency data: An application to forecasting US GNP at monthly intervals," *EconomPaper*, No. 90-5, 1990.
3. Ghysels E., Santa-Clara P., and Valkanov R., "The MIDAS touch: Mixed data sampling regression models," 2004.
4. Ghysels E., "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, Vol. 193, No. 2, 2016. pp. 294-314.
5. Asimakopoulous S., Paredes J., and Warmedinger T., "Forecasting fiscal time series using mixed frequency data," *ECB Working Paper*, No. 1550, 2013.
6. Viefers P. Bayesian inference for the mixed-frequency VAR model. 2011.
7. Mariano R., Ozmucur S. Predictive Performance of Mixed-Frequency Nowcasting and Forecasting Models (with Application to Philippine Inflation and GDP Growth). 2020.
8. Diebold F., "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests," *Journal of Business & Economic Statistics* 33, Vol. 33, No. 1, 2015. pp. 1-1.
9. Mariano R., Preve D., "Statistical tests for multiple forecast comparison," *Journal of econometrics*, Vol. 169, No. 1, 2012. pp. 123-130.
10. Fair R., Shiller R., "The informational content of ex ante forecasts," *Review of Economics and Statistics*, No. 71, 1989. pp. 325-331.
11. Chong Y., Hendry D., "Econometric evaluation of linear macro-economic models," *Review of Economic Studies*, No. 53, 1986. pp. 671-690.
12. Кудрявцева И., "Анализ эффективности расширенного фильтра Калмана, сигма-точечного фильтра Калмана и сигма-точечного фильтра частиц," *Научный вестник Московского государственного технического университета гражданской авиации*, Vol. 2, No. 224, 2016.
13. Chen B., Zdrozny P., "An extended Yule-Walker method for estimating a vector autoregressive model with mixed-frequency data," *Advances in Econometrics*, No. 13, 1998. pp. 47-74.
14. Ghysels E., Santa-Clara P., and Valka R., "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, Vol. 131, No. 1-2, 2006. pp. 59-95.

15. Bai J., Ghysels E., and Wright J., "State space models and MIDAS regressions," *Econometric Reviews*, Vol. 32, No. 7, 2013. pp. 779-813.
16. Clements M., Galvão A., "Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States," *Journal of Business & Economic Statistics*, Vol. 26, No. 4, 2008. pp. 546-554.
17. Andreou E., Ghysels E., and Kourtello A., "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business & Economic Statistics*, Vol. 31, No. 2, 2013. pp. 240-251.
18. Ghysels E., Qian H., "Estimating MIDAS regressions via OLS with polynomial parameter profiling," *Econometrics and statistics*, Vol. 9, 2019. pp. 1-16.
19. Cole S., Chu H., and Greenlan S., "Maximum likelihood, profile likelihood, and penalized likelihood: a primer," *American journal of epidemiology*, Vol. 179, No. 2, 2014. pp. 252-260.
20. Smith R., Giles D. *The Almon estimator: Methodology and users' guide*. Reserve Bank of New Zealand, 1976.
21. Ghysels E., Sinko A., and Valkanov R., "MIDAS regressions: Further results and new directions," *Econometric reviews*, Vol. 26, No. 1, 2007. pp. 53-90.
22. Clement M. *Variable selection in predictive MIDAS models*. 2014.
23. Siliverstovs B., "Short-term forecasting with mixed-frequency data: a MIDASSO approach," *Applied Economics*, Vol. 49, No. 13, 2017. pp. 1326-1343.
24. Ghysels E et al., "MIDAS-PRO-LASSO: Mixed Frequency DataRegression Models with Parameter Profiling and LASSO," *Discussion Paper LFin - UCL Louvain-la-Neuve* 2017.
25. Foroni C et al., "Forecasting daily electricity prices with monthly macroeconomic variables," 2019.
26. Gneiting T and Raftery A, "Strictly proper scoring rules, prediction, and estimation," *technical Report no. 463*, Department of Statistics, University of Washington, Seattle, Washington, USA 2004.
27. Christoffersen P., Diebold F., "Financial asset returns, direction-of-change forecasting and volatility dynamics," *Management Science*, No. 52, 2006. pp. 1237-1287.
28. Hansen P., Lunde A., and Nason J., "The Model Confidence Set," *Econometrica*, No. 79, 2011. pp. 453-497.
29. Abel R., Puggioni G., "Mixed frequency models: Bayesian approaches to estimation and prediction," *International Journal of Forecasting*, Vol. 26, No. 2, 2010. pp. 293-311.
30. Lindley D., "A statistical paradox," *Biometrika*, Vol. 44, No. 1/2, 1957. pp. 187-192.

31. Spanos A., "Who should be afraid of the Jeffreys-Lindley paradox?," *Philosophy of Science*, Vol. 80, No. 1, 2013. pp. 73-93.
32. Uhlig H., "Bayesian vector autoregressions with stochastic volatility," *Journal of the Econometric Society*, 1997. pp. 59-73.
33. Brave S., Butters R., and Justiniano A. *Forecasting economic activity with mixed frequency Bayesian VARs*. 2016.
34. Schorfheide F., Song D., "Real-time forecasting with a mixed-frequency VAR," *Journal of Business & Economic Statistics*, Vol. 33, No. 3, 2015. pp. 366-380.
35. Svitak J. *Nowcasting with many predictors: Predicting U.S. GDP Using Bayesian-Structural Time-Series Model*. University of Amsterdam, 2014.
36. Harvey A. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.
37. Koopman S., "Disturbance smoother for state space models," *Biometrika*, Vol. 80, No. 1, 1993. pp. 117–126.
38. Weron R. *Electricity price forecasting: A review of the state-of-the-art with a look into the future* // *International journal of forecasting*. – 2014. – T. 30. – №. 4. – C. 1030-1081.
39. Chen Y. et al. *Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings* // *Applied Energy*. – 2017. – T. 195. – C. 659-670.
40. Deng Z., Liu C., Zhu Z. *Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network* // *International Journal of Electrical Power & Energy Systems*. – 2021. – T.,.
41. Guo X. et al. *A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price* // *Energy Reports*. – 2020. – T. 6. – C. 1046-1053.
42. Kim T. Y., Cho S. B. *Predicting residential energy consumption using CNN-LSTM neural networks* // *Energy*. – 2019. – T. 182. – C. 72-81.
43. Kuo P. H., Huang C. J. *A high precision artificial neural networks model for short-term energy load forecasting* // *Energies*. – 2018. – T. 11. – №. 1. – C. 213.
44. Qu Z. et al. *Probability Prediction Method of Short-Term Electricity Price Based on Quantile Neural Network Model* // *Journal of Electrical Engineering & Technology*. – 2020. – T. 15. – №. 2. – C. 547-559.
45. Ugurlu U., Oksuz I., Tas O. *Electricity price forecasting using recurrent neural networks* // *Energies*. – 2018. – T. 11. – №. 5. – C. 1255.

46. Cont R. Empirical properties of asset returns: stylized facts and statistical issues. – 2001.
47. Oord A. et al. Wavenet: A generative model for raw audio //arXiv preprint arXiv:1609.03499. – 2016.
48. LeCun Y. et al. Convolutional networks for images, speech, and time series //The handbook of brain theory and neural networks. – 1995. – Т. 3361. – №. 10. – С. 1995.
49. Чучуева И. А. Модель прогнозирования временных рядов по выборке максимального подобия //Москва. – 2012.
50. Быков Ф. Л., Гордин В. А. Краткосрочный прогноз часового потребления электроэнергии с учетом погоды для субъектов РФ //Известия РАН: Энергетика. – 2017. – №. 5. – С. 47-56.
51. Tian C. et al. Data driven parallel prediction of building energy consumption using generative adversarial nets //Energy and Buildings. – 2019. – Т. 186. – С. 230-243.
52. Cuaresma J. C. et al. Forecasting electricity spot-prices using linear univariate time-series models //Applied Energy. – 2004. – Т. 77. – №. 1. – С. 87-106.
53. Fumo N., Biswas M. A. R. Regression analysis for prediction of residential energy consumption //Renewable and sustainable energy reviews. – 2015. – Т. 47. – С. 332-343.
54. Roldán-Blay C. et al. Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model //Energy and Buildings. – 2013. – Т. 60. – С. 38-46.
55. TOROS H., Aydın D. Prediction of Short-Term Electricity Consumption by Artificial Neural Networks Using Temperature Variables //Avrupa Bilim ve Teknoloji Dergisi. – 2018. – №. 14. – С. 393-398.
56. Carmona R., Coulon M. A survey of commodity markets and structural models for electricity prices //Quantitative Energy Finance. – Springer, New York, NY, 2014. – С. 41-83.
57. Sainath T. N. et al. Convolutional, long short-term memory fully connected deep neural networks //2015 IEEE international conference on acoustics, speech, and signal processing (ICASSP). – IEEE, 2015. – С. 4580-4584.
58. Ullah A. et al. Action recognition in video sequences using deep bi-directional LSTM with CNN features //IEEE access. – 2017. – Т. 6. – С. 1155-1166.
59. Oh S. L. et al. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats //Computers in biology and medicine. – 2018. – Т. 102. – С. 278-287.
60. Zhao R. et al. Learning to monitor machine health with convolutional bi-directional LSTM networks //Sensors. – 2017. – Т. 17. – №. 2. – С. 273.

61. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate //arXiv preprint arXiv:1409.0473. – 2014.
62. Borovykh A., Bohte S., Oosterlee C. W. Conditional time series forecasting with convolutional neural networks //arXiv preprint arXiv:1703.04691. – 2017.
63. Nguyen H. T., Nabney I. T. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models //Energy. – 2010. – Т. 35. – №. 9. – С. 3674-3685.
64. Memarzadeh G., Keynia F. Short-term electricity load and price forecasting by a new optimal LSTM-NN based prediction algorithm //Electric Power Systems Research. – 2021.
65. TowardsDatascience [Электронный ресурс] [2020]. URL: <https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9>.
66. АТСЭнерго [Электронный ресурс] [2021]. URL: <https://www.atsenergo.ru/>.
67. Cleveland R. B. et al. STL: A seasonal-trend decomposition //Journal of official statistics. – 1990. – Т. 6. – №. 1. – С. 3-73.
68. Kostrzewski M.A.K.J., "Probabilistic electricity price forecasting with Bayesian stochastic volatility models. ," Energy Economics, No. 80, 2019. pp. 610-620.

В СЕРИИ ПРЕПРИНТОВ
РАНХиГС РАССМАТРИВАЮТСЯ
ТЕОРЕТИЧЕСКИЕ
И ПРАКТИЧЕСКИЕ ПОДХОДЫ
К СОЗДАНИЮ, АКТИВНОМУ
ИСПОЛЬЗОВАНИЮ
ВОЗМОЖНОСТЕЙ
ИННОВАЦИЙ В РАЗЛИЧНЫХ
СФЕРАХ ЭКОНОМИКИ
КАК КЛЮЧЕВОГО УСЛОВИЯ
ЭФФЕКТИВНОГО УПРАВЛЕНИЯ



РАНХиГС
РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ