

**Федеральное государственное бюджетное образовательное учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА и
ГОСУДАРСТВЕННОЙ СЛУЖБЫ
при Президенте Российской Федерации»**

**ИНСТИТУТ ЭКОНОМИКИ, МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ
ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ
НАПРАВЛЕНИЕ 38.03.01 ЭКОНОМИКА**

Группа ЭО-15-01

Кафедра микроэкономики

Допустить к защите
заведующий кафедрой микроэкономики

_____ М.И. Левин

«____» _____ 201__ г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**ПРОГНОЗИРОВАНИЕ
ИЕРАРХИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ**

студент-бакалавр
Касьянова Ксения Алексеевна

/_____ /_____/
(подпись) (дата)

научный руководитель выпускной
квалификационной работы
ст. преп. Демешев Борис Борисович

/_____ /_____/
(подпись) (дата)

**МОСКВА
2019 г.**

Оглавление

Введение	3
1 Модели прогнозирования временных рядов с иерархической структурой	4
1.1 Обзор литературы	4
1.2 Комбинирование прогнозов	5
1.2.1 Метод «снизу-вверх»	6
1.2.2 Метод «сверху-вниз»	7
1.2.3 Оптимальная комбинация по OLS	7
2 Сравнение моделей прогнозирования	9
2.1 Описание данных	9
2.1.1 Квартальные данные	10
2.1.2 Квартальные сезонно сплаженные данные.	11
2.1.3 Месячные данные	11
2.2 Выбор моделей прогнозирования рядов нижнего уровня	12
2.3 Группировка рядов третьего уровня по регионам, по типу и по кластерам	15
2.4 Сравнение иерархических моделей.	17
Заключение	25
Список литературы	26
Приложение А Коды для выгрузки данных	28
Приложение Б Визуализация временных рядов с трехуровневой структурой	30
Приложение В Сравнение иерархических моделей	32

Введение

В анализе данных часто встречаются данные со сложной многоуровневой структурой, точный прогноз которых является одним из ключевых факторов принятия эффективных решений. Необходимость прогнозирования данных с иерархической структурой возникает, например, в микроэкономике при анализе спроса, макроэкономике, страховании, демографии и т.д.

В данной работе исследуются методы прогнозирования иерархических временных рядов, учитывающие зависимость между уровнями агрегирования и внутри одного уровня. Цель работы: сравнение моделей, учитывающих иерархическую структуру данных, выявление факторов, позволяющих улучшить прогнозы агрегированного временного ряда. Достижение поставленной цели предполагает постановку и решение следующих задач:

- сбор данных с трехуровневой иерархической структурой;
- выбор моделей для прогнозирования агрегированного ряда;
- сравнение различных методов комбинирования прогнозов нижних рядов, полученных по выбранным моделям;
- сравнение качества прогнозов полученным по рядам второго уровня, сгруппированным по оригинальным (по территориальному признаку и по классу) и полученным с помощью кластеризации признакам;
- прогнозирование рядов второго и третьего уровня по выбранным моделям, сравнение суммы и оптимальной комбинации этих прогнозов с прогнозом агрегированного временного ряда.

Применение иерархических моделей на трех наборах данных, обладающих различными характеристиками, с использованием кросс-валидации со скользящим окном позволяет получить более устойчивые выводы.

Данная работа состоит из введения, двух глав основной части, заключения и приложений. В первой главе рассматриваются основные модели прогнозирования иерархических временных рядов. Во второй главе проводится сравнение моделей применительно к собранным данным с требуемой структурой. В приложении (А) содержатся коды для выгрузки данных. В приложении (Б) содержатся графики, отражающие многоуровневую структуру данных. В приложении (В) представлены таблицы, позволяющие сравнить качество прогнозов, полученное по различным моделям.

1 Модели прогнозирования временных рядов с иерархической структурой

1.1 Обзор литературы

В современной литературе по анализу данных можно выделить несколько основных подходов к прогнозированию временных рядов с иерархической структурой. Основное предположение, на котором построены многие из методов, заключается в том, что при группировке рядов, которые ведут себя одинаково, характерные ошибки внутри групп будут компенсировать друг друга, в то время как более релевантная для прогнозирования индивидуальная динамика будет сохраняться.

Одним из способов повышения точности прогнозов является агрегирование данных до составления прогнозов, другой – комбинирование самих прогнозов. Информация полученная из агрегированных рядов может иметь существенное влияние при прогнозировании рядов нижнего уровня, хотя ее использование может сопровождаться некоторыми сложностями, возникающими из-за индивидуальных особенностей рядов.

Принципиально другим методом работы с иерархическими временными рядами является байесовский подход. Для рядов относящихся к одной группе подбирается одинаковое априорное распределение коэффициентов. В простых случаях эффект от выбора такого априорного распределения такой: если ряды сильно отличаются от сгруппированных, то влияние группового априорного распределения уменьшается, апостериорное распределение коэффициентов для них сужается.

Альтернативными моделями, подходящими для прогнозирования временных рядов с многоуровневой структурой являются, например, модель векторной авторегрессии (VAR), в которой временные ряды, относящиеся к одной группе, могут иметь общие параметры. В качестве некой альтернативы этому методу можно предложить использование модели ARIMA с дополнительными регрессорами, полученными из прогнозируемого набора данных.

Для прогнозирования иерархических рядов в том числе применяются многомерные модели пространства состояний, векторное экспоненциальное сглаживание, а также байесовские подходы применительно к пулу аналогичных временных рядов[8]. Эмпирические результаты показали, что с помощью перечисленных выше методов точность прогноза может быть улучшена, поскольку при оценке параметров учитывается зависимость между временными рядами, принадлежащими одной группе. Однако использование этих моделей связано с выполнением большого числа предпосылок или введения соответствующих ограничений.

Во многих теоретических и эмпирических работах было замечено, что зачастую более простые методы оказываются в разы эффективнее, сложных моделей с большим чис-

лом оцениемых коэффициентов. Такой подход к прогнозированию рядов, как получение оптимальной комбинации прогнозов по качеству прогнозов может легко обогнать модели, учитывающие особенности каждого из рядов при оценивании коэффициентов.

1.2 Комбинирование прогнозов

Существует несколько популярных подходов к комбинированию прогнозов: «сверху-вниз», «снизу-вверх», «из середины» и корректировка прогнозов по OLS. Данные, к которым эти подходы применимы, отличаются следующим свойством: когда ряды нижнего уровня группируются, прогнозы каждой группы должны быть равны сумме прогнозов рядов, входящих в группу.

Для любого набора данных, обладающего таким свойством можно построить суммирующую матрицу S , которая отражает структуру иерархии рядов. Примером такой матрицы является:

$$\begin{bmatrix} y_t \\ y_{1,t} \\ y_{2,t} \\ \dots \\ y_{i,t} \\ y_{11,t} \\ y_{12,t} \\ y_{13,t} \\ y_{21,t} \\ y_{22,t} \\ y_{23,t} \\ \dots \\ y_{ij-2,t} \\ y_{ij-1,t} \\ y_{ij,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{11,t} \\ y_{12,t} \\ y_{13,t} \\ y_{21,t} \\ y_{22,t} \\ y_{23,t} \\ \dots \\ y_{ij-2,t} \\ y_{ij-1,t} \\ y_{ij,t} \end{bmatrix}, \quad (1.1)$$

где $y_{ij,t}$, $y_{i,t}$, y_t – значения j -го ряда третьего уровня, i -го ряда второго уровня и ряда первого уровня соответственно в момент времени t . Схематически такая структура данных представлена на рисунке (1.1).

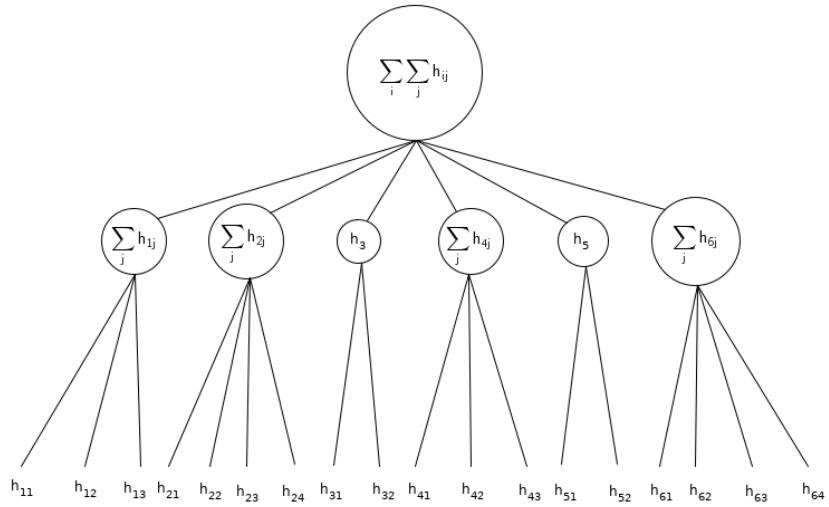


Рисунок 1.1 — Структура временных рядов, необходимых для исследования

Переписав уравнение (1.1) в более компактной форме имеем:

$$y_t = Sb_t, \quad (1.2)$$

где y_t – вектор размерности $(n \times 1)$ всех наблюдений на всех уровнях иерархии в момент времени t , $n = 1 + i + \sum_j j$, S – суммирующая матрица, отражающая иерархическую структуру данных, b_t – вектор размерности $(m \times 1)$ всех наблюдений на самом нижнем уровне иерархии в момент времени t , $m = \sum_j j$

1.2.1 Метод «снизу-вверх»

Самый распространенный метод получения прогнозов для всех уровней иерархии – «снизу-вверх» метод (англ. bottom-up). Суть метода заключается в получении независимых прогнозов на h шагов вперед для рядов нижнего уровня иерархии и их агрегирование согласно структуре иерархии S :

$$\tilde{y}_h = S\hat{y}_{K,h}, \quad (1.3)$$

где \tilde{y}_h – собранные с помощью суммирования прогнозы рядов уровней $1 \dots K - 1$ и базовые прогнозы $\hat{y}_{K,h}$.

Основное преимущество этого подхода – при таком методе агрегации не теряется никакая информация. Недостаток метода заключается в том, что ряды нижнего уровня могут быть сильно зашумленными, а потому более сложными для моделирования и прогнозирования [23]. В таком случае метод «сверху-вниз» может быть эффективнее., но это зависит от того, насколько точными являются прогнозы агрегированного ряда и подобранные пропорции.

1.2.2 Метод «сверху-вниз»

Аналогичный предыдущему подход – метод «сверху-вниз» (англ. top-down). Для этого метода необходимо получить базовый прогноз для агрегированного ряда (первого уровня), а также коэффициенты-пропорции на которые агрегированный ряд будет умножаться для получения прогноза нижнего уровня.

Прогноз агрегированного ряда разбивается на прогнозы компонент по весам, полученным на основе исторических данных. Например, это могут быть усредненные за весь период доли i -ого ряда в агрегированном. Также веса могут оцениваться с учетом веса прогноза i -ого ряда в прогнозе агрегированного.

Также похожим подходом является метод «из середины» (англ. middle-out), объединяющий в себе два предыдущих метода. В этом случае сначала генерируются прогнозы для среднего уровня, а на их основе согласно методам описанным выше получаются прогнозы для верхних и нижних уровней.

1.2.3 Оптимальная комбинация по OLS

Методы «снизу-вверх» и «сверху-вниз», не учитывают корреляцию между рядами на каждом уровне. Один из методов прогнозирования, позволяющий сделать это, был предложен Хайндманом и др.[18]. Этот метод позволяет получать скорректированные оценки прогнозов агрегированных рядов и их компонент. Основная идея этого метода заключается в получении согласованных на разных уровнях прогнозов на основе независимых между собой прогнозов всех рядов всех уровней.

Подход, описанный Хайндманом[13] предполагает получение прогноза для каждого из рядов иерархии. Предположение следующее: поскольку эти базовые прогнозы генерируются независимо друг от друга, они не могут обладать свойством аддитивности, то есть они не будут суммироваться в соответствии с иерархической структурой, отраженной в матрице S . Данный подход позволяет получить оптимальную комбинацию прогнозов, полученных на основе несогласованных базовых прогнозов. Прогнозы для каждого ряда переоцениваются так, чтобы они были максимально приближены к базовым прогнозам, но в то же время удовлетворяли иерархической структуре S .

Идея заключается в представлении базовых прогнозов на h шагов вперед в рамках линейной регрессии:

$$\hat{y}_h = S\beta_h + e_h, \quad (1.4)$$

где \hat{y}_h – вектор базовых прогнозов на h шагов вперед для всех уровней иерархии, S – матрица суммирования, $\beta_h = E[\beta_{n+h}|y_1, \dots, y_n]$ – неизвестное среднее значение будущих значений

рядов нижнего уровня, $e_h \sim N(0, \Sigma_h)$ – ошибка регрессии. В общем случае Σ_h неизвестна. Однако можно показать, что при расчете точечных прогнозов корреляция между рядами не важна.

Если базовые прогнозы приблизительно удовлетворяют иерархической структуре S , то и ошибки также должны приблизительно удовлетворять структуре иерархической агрегации:

$$\hat{e}_h \approx S e_{K,h}, \quad (1.5)$$

где $e_{K,h}$ – вектор ошибок для рядов нижнего уровня. Такое условие должно выполняться для любого разумного набора прогнозов. При этом предположении, лучшей линейной несмещенной оценкой для β_h является:

$$\hat{\beta}_h = (S'S)^{-1}S'\hat{y}_h. \quad (1.6)$$

Получив оценки $\hat{\beta}_h$ можно пересчитать скорректированные прогнозы для всех уровней иерархии:

$$\tilde{y}_h = S\hat{\beta}_h = S(S'S)^{-1}S'\hat{y}_h. \quad (1.7)$$

Согласно полученной формуле 1.7 пересчитываются скорректированные прогнозы, не зависящие от Σ_h . Для получения интервальных оценок считается $Var(\tilde{y}_h)$:

$$Var(\tilde{y}_h) = S(S'\Sigma_h^{-1}S)^{-1}S. \quad (1.8)$$

Преимуществом этого метода является то, что этот подход использует доступную в иерархии информацию, учитывает взаимосвязь между прогнозами на всех уровнях иерархии и при условии, что базовые прогнозы являются несмещенными, скорректированные прогнозы также будут несшенными.

2 Сравнение моделей прогнозирования

2.1 Описание данных

Для анализа необходимо найти наборы данных удовлетворяющие следующим критериям: структура трехуровневая и иерархическая, обладающая свойством аддитивности, т.е. для I рядов второго уровня, каждый из которых делится на J рядов третьего уровня, выполняется:

$$y_t = \sum_{i=1}^I y_{i,t} = \sum_{i=1}^I \sum_{j=1}^J y_{ij,t}, \quad (2.1)$$

где $y_{ij,t}$, $y_{i,t}$, y_t – значения j -го ряда третьего уровня, i -го ряда второго уровня и ряда первого уровня соответственно в момент времени t .

Методы иерархического прогнозирования позволяют суммировать прогнозы на каждом уровне, чтобы получить прогнозы на уровне выше. Если данные группируются, прогнозы для агрегированного ряда по группе должны быть равны сумме прогнозов отдельных рядов, входящих в группу.

Поиск реальных данных, идеально подходящих под такую структуру, затруднен. Для многих микроэкономических показателей в первую очередь собираются данные по отдельным компонентам, из которых можно получить агрегированные ряды, удовлетворяющие свойству аддитивности. Поскольку получить такие данные сложно, воспользуемся доступными макроэкономическими данными. При их использовании необходимо учитывать, что значение верхнего ряда не всегда в точности равняется сумме нижних рядов. Причиной этому могут быть различия в методологиях расчета показателей рядов разных уровней.

Так например, разбивая ряд ВВП на компоненты по регионам и отраслям, надо учесть, что они будут отражать несколько иной показатель – валовую добавленную стоимость (ВДС)¹. Агрегированный ряд, получаемый при суммировании всех ВДС, будет меньше ВВП на величину чистых субсидий на производство и импорт. Такой показатель имеет близкую к единице корреляцию с рядом ВВП, поэтому при точном его прогнозировании мы можем получить представление как об общей динамике всех компонент, составляющих ряд, так и о динамике ряда ВВП. Так как целью работы является сравнение моделей, для упрощения будем работать с агрегированными показателями по ВВП, являющиеся простой суммой из рядов нижнего уровня.

Этот факт также учитывается при расчете вклада компонент, составляющих ряд, в

¹ Валовая добавленная стоимость определяется как разность между выпуском товаров и услуг и их промежуточным потреблением. ВДС исчисляется на уровне отраслей и отражает образование первичных доходов в результате процесса производства товаров и услуг.

процентное изменение агрегированного показателя, не обладающего свойством аддитивности [9]:

$$C\%\Delta_{i,t} = 100 * \frac{q_{i,t} - q_{i,t-1}}{\sum_j q_{j,t-1}}, \quad (2.2)$$

где $q_{i,t}$ – значение i -ого ряда в момент времени t . Такой показатель позволяет определить изменения в структуре агрегата, что делает его ценным инструментом экономического анализа. Если при прогнозировании с помощью иерархических моделей удастся улучшить прогноз агрегированного ряда, то фактически мы также сможем получить достаточно точные прогнозы показателей вклада каждой компоненты.

Для анализа были выбраны три набора данных с описанными выше свойствами, обладающие разной сезонностью: квартальные, квартальные сезонно сглаженные и месячные данные. В следующих пунктах будут более подробно описаны особенности каждого из наборов данных. Ознакомиться с визуальной представлением этих наборов можно в приложении (Б).

2.1.1 Квартальные данные

Квартальные данные² – ряды ВДС по 28 странам Европейского союза (включая Великобританию) в разбивке по основным отраслям³:

- А – сельское хозяйство, лесное хозяйство и рыболовство;
- В – промышленность (кроме строительства);
- F – строительство;
- G – оптовая и розничная торговля, транспорт, услуги общественного питания и т.д.;
- J – информация и связь;
- К – финансовая и страховая деятельность;
- L – операции с недвижимостью;
- М – профессиональная, научно-техническая, административная деятельность;
- О – государственное управление, оборона, образование, здравоохранение и социальная работа;
- Р – искусство, развлечения, отдых и другие виды услуг.

Данные выгружены за период с 2000-Q1 по 2018-Q3. Разница между совокупным ВВП всех 28 стран, входящих в состав ЕС и суммой ВДС по всем отраслям для каждого из государств, не превышает 1.5% от ВВП.

²Eurostat: European statistics – Database. URL: <https://ec.europa.eu/eurostat/data/database> (дата обращения: 16.02.2019). Код для выгрузки данных с сайта в приложении (А.1).

³Eurostat metadata: Annual national accounts (nama10). URL: https://ec.europa.eu/eurostat/cache/metadata/en/nama10_esms.htm (дата обращения: 16.02.2019).

2.1.2 Квартальные сезонно сглаженные данные

Квартальные сезонно сглаженные данные⁴ – это ряды ВДС для каждого из 50 штатов Америки с разбивкой на 21 отрасль. Данные выгружены за период с 2005-Q1 по 2018-Q2. В этом наборе 11 рядов имели пропуски. По четырем из этих рядов данные перестали собираться в 2008 году, поэтому эти ряды были исключены целиком. Остальные пропуски были заполнены с помощью экспоненциально взвешенного скользящего среднего с шириной окна 4⁵.

Квартальные оценки ВДС в США пересчитываются с учетом сезонных колебаний следующим образом: BEA⁶ оценивает соответствующие коэффициенты сезонной корректировки, после чего удаляет из временного ряда среднее влияние изменений, которые обычно происходят примерно в одно и то же время с одинаковой величиной каждый год. Сезонно несглаженные ряды по этому показателю BEA не публикует.

Показатели по ВДС публикуются в реальном денежном эквиваленте (за базовый год принимается 2012). Значения реальных показателей ВДС по отраслям не обязательно дают в сумме показатель реального ВДС для каждого штата за интересующий период, поскольку относительные цены, используемые в качестве весов для корректировки показателей по отраслям, отличаются от общего уровня цен используемых для корректировки агрегированного показателя. Эта разница не превышает 0.5% ВВП. Для периодов близких к 2012 году, когда значительных отклонений относительных цен от индекса цен по стране не было, показатель ВДС штата совпадает с суммой ВДС по отраслям. Разница между ВВП США и суммой ВДС по отраслям для каждого штата не превышает 2%.

2.1.3 Месячные данные

Месячные данные – показатели рождаемости⁷ и смертности⁸ по основным причинам в каждом регионе РФ, дающие в сумме естественный прирост населения помесячно. Данные выгружены за период с 2006-01 по 2019-01.

Если для каждого из регионов все показатели из набора данных «Число зарегистрированных умерших по основным классам и отдельным причинам смерти» просуммировать

⁴FRED: Economic Data. URL: <https://fred.stlouisfed.org/> (дата обращения: 18.02.2019). Код для выгрузки данных с сайта в приложении (A.2).

⁵Алгоритм, используемый в пакете R imputeTS имеет адаптивный размер окна: в случае длинных промежутков с пропущенными значениями, размер окна постепенно увеличивается до тех пор, пока не появятся как минимум 2 значения не-NA.

⁶U.S. Bureau of Economic Analysis (BEA). URL: <https://www.bea.gov/> (дата обращения: 25.02.2019)

⁷ЕМИСС: государственная статистика. Число зарегистрированных родившихся (оперативные данные). URL: <https://fedstat.ru/indicator/33555> (дата обращения: 10.04.2019).

⁸ЕМИСС: государственная статистика. Число зарегистрированных умерших по основным классам и отдельным причинам смерти (оперативные данные). URL: <https://fedstat.ru/indicator/33559> (дата обращения: 10.04.2019).

по причинам смерти, значения будут отличаться от показателей из набора данных «Число зарегистрированных умерших». Такое расхождение объясняется тем, что для первого набора разрабатываются ряды только по основным классам и отдельным причинам смерти, имеющим наибольший вес. Также в 2011 году методика разработки показателя была пересмотрена, чтобы соответствовать международной статистической классификации⁹.

Для анализа необходимы ряды, в сумме дающие агрегированный ряд естественного прироста населения. Для каждого региона имеем следующее разбиение:

- РО – число рожденных;
- УБ – число умерших из-за болезней (болезней органов дыхания, органов пищеварения, системы кровообращения, инфекционных и паразитарных болезней, новообразований);
- УУ – число умерших по причине убийства и самоубийства;
- УВ – число умерших по прочим причинам (отравление алкоголем, транспортные травмы всех видов и внешние причины).

Три группы причин смертности были выделены согласно международной статистической классификации, чтобы избежать большого числа рядов с нулевыми и близкими к нулю значениями. Разница между показателем смертности по каждому региону и суммой по всем причинам смертности была добавлена к ряду «УВ».

С 2015 года также собираются данные по Республике Крым и городу федерального значения Севастополю. Однако данных нужной сезонности по каждому из классов за 2006-2014 годы Держстат Украины¹⁰ не предоставляет, поэтому ряды по этим регионам были исключены из набора данных.

2.2 Выбор моделей прогнозирования рядов нижнего уровня

Для того чтобы определить, можно ли с помощью комбинирования прогнозов получить более точные прогнозы агрегированных рядов необходимо выбрать модель для прогнозирования нижних рядов. Конечно, для каждого ряда можно подобрать свою модель с параметрами, оптимизирующими, например, метрику качества прогноза. Недостатками такого метода являются зависимость от выбора метрики, неинтерпретируемость результатов и вычислительная сложность алгоритма, перебирающего все модели для большого числа рядов. Использование одной и той же модели для всех рядов иерархии позволит увидеть, есть ли зависимость между выбором параметров модели и методом комбинирования рядов или прогнозов.

⁹Демографический ежегодник России: методические пояснения. URL: http://www.gks.ru/bgd/regl/B17_16/Main.htm (дата обращения: 09.05.2019).

¹⁰Держстат Украины – Статистика населения Украины. URL: http://database.ukrcensus.gov.ua/MULT/Dialog/statfile_c.asp (дата обращения: 12.05.2019).

Эффективность каждого из методов комбинирования будет проверяться на десяти различных моделях: AR с малым числом лагов (с линейным трендом), AR с линейным и с квадратичным трендом, интегрированная AR, ARMA с линейным трендом, ARIMA, ETS с фиксированными параметрами, ARIMA, ETS и TBATS с автоматическим подбором параметров.¹¹

Для выбора параметров в моделях применяется кросс-валидация со скользящим окном с шагом в одно наблюдение. К рядам нижнего уровня будет применяться модель, для которой среднее по всем подвыборкам RMSE, полученное на кросс-валидации для агрегированного ряда, будет ниже других в классе используемой модели.

Ширина окна для каждого из наборов данных подбиралась с учетом длины ряда и горизонта прогнозирования в два года таким образом, чтобы при проведении кросс-валидации с шагом в один год получалось не менее пяти подвыборок. В результате

- для квартальных рядов по ВВП ЕС ширина окна 48, прогноз на 8 шагов вперед;
- для сезонно сглаженных рядов по ВВП США ширина окна 28, прогноз на 8 шагов вперед;
- для месячных рядов по естественному приросту РФ ширина окна 84, прогноз на 24 шага вперед.

Результат перебора параметров для каждой из основных моделей для всех трех наборов данных представлен в таблице (2.1). Для моделей класса ARIMA число лагов в AR части p подбиралось среди значений меньших или равных периодичности рядов: для квартальных – 4, для месячных – 12, число лагов в MA части q – меньшее либо равное двум. В ETS модели перебирались все возможные варианты параметров, в TBATS – тренд мог быть демпфированным, либо нет. Для всех моделей, параметр λ , отвечающий за преобразование Бокса-Кокса, либо равнялся 1, либо оценивался в модели.

Для сравнения качества прогнозов будут использоваться следующие метрики:

- средняя ошибка (mean error):

$$ME = \frac{1}{h} \sum_{i=1}^h (\hat{y}_{t+i|t} - y_{t+i}); \quad (2.3)$$

- квадратный корень из среднеквадратичной ошибки (root mean square error):

$$RMSE = \sqrt{\frac{1}{h} \sum_{i=1}^h (\hat{y}_{t+i|t} - y_{t+i})^2}; \quad (2.4)$$

¹¹Автоматический перебор параметров модели осуществляется с помощью функций R `auto.arima`, `ets` и `tbats` соответственно.

Таблица 2.1 — Параметры моделей

	Квартальные	Сезонно сглаженные	Месячные
AR с линейным трендом (с малым числом лагов)	$(p, d, q) = (2, 0, 0)$, $(P, D, Q)_4 = (1, 0, 0)$	$(p, d, q) = (2, 0, 0)$	$(p, d, q) = (2, 0, 0)$, $(P, D, Q)_{12} = (1, 0, 0)$
AR с линейным трендом	$(p, d, q) = (3, 0, 0)$, $(P, D, Q)_4 = (2, 0, 0)$	$(p, d, q) = (4, 0, 0)$	$(p, d, q) = (11, 0, 0)$, $(P, D, Q)_{12} = (2, 0, 0)$
AR с квадратичным трендом	$(p, d, q) = (3, 0, 0)$, $(P, D, Q)_4 = (2, 0, 0)$	$(p, d, q) = (4, 0, 0)$	$(p, d, q) = (11, 0, 0)$, $(P, D, Q)_{12} = (2, 0, 0)$
Интегрированная AR	$(p, d, q) = (3, 1, 0)$, $(P, D, Q)_4 = (2, 1, 0)$	$(p, d, q) = (4, 1, 0)$	$(p, d, q) = (4, 0, 0)$, $(P, D, Q)_{12} = (1, 1, 0)$
ARMA с линейным трендом	$(p, d, q) = (3, 0, 1)$, $(P, D, Q)_4 = (2, 0, 1)$	$(p, d, q) = (4, 0, 1)$	$(p, d, q) = (4, 0, 1)$, $(P, D, Q)_{12} = (1, 0, 1)$
ARIMA	$(p, d, q) = (3, 1, 1)$, $(P, D, Q)_4 = (2, 1, 1)$	$(p, d, q) = (4, 1, 1)$	$(p, d, q) = (4, 1, 1)$, $(P, D, Q)_{12} = (1, 1, 1)$
ARIMA с автоматическим подбором параметров	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$
ETS с фиксированными параметрами	$(E, T, S) = (M, M, M)$ $\lambda = 1$	$(E, T, S) = (A, A, A)$ $\lambda = 1$	$(E, T, S) = (A, Ad, A)$ $\lambda = 1$
ETS с автоматическим подбором параметров	$\lambda = 1$	$\lambda = 1$	$\lambda = 1$
TBATS	$\lambda = 1, T = A$	$\lambda = 1, T = A$	$\lambda = 1, T = Ad$

(p, d, q) – параметры для ARIMA модели, $(P, D, Q)_m$ – параметры для SARIMA модели с периодичностью m , λ – параметр преобразования Бокса-Кокса, T - параметр, отвечающий за характер тренда (A - аддитивный, Ad - аддитивный демпфированный, M - мультипликативный)

– средняя абсолютная ошибка в процентах (mean absolute percentage error):

$$MAPE = \frac{1}{h} \sum_{i=1}^h \frac{|y_{t+i} - \hat{y}_{t+i|t}|}{y_{t+i}} * 100\%. \quad (2.5)$$

Основной метрикой сравнения точности прогнозов моделей будет RMSE. Средняя ошибка (ME) позволит понять, насколько хорошо модель улавливает тренд в рядах. MAPE в качестве метрики сравнения точности прогнозов является смещенным показателем, поскольку он будет систематически выбирать модель, прогнозы которой занижены, так как на отрицательные ошибки налагаются большие штрафы, чем на положительные. Но зато MAPE позволит сравнивать улучшение качества прогнозов для разных наборов данных, хотя для набора данных по России MAPE не является показательной метрикой, так как в нем имеются нулевые и близкие к нулю значения.

2.3 Группировка рядов третьего уровня по регионам, по типу и по кластерам

Для данного исследования подбирались наборы данных с трехуровневой структурой, обладающие свойством аддитивности. Это позволяет проверить, можно ли улучшить прогноз агрегированного ряда используя комбинации прогнозов каждого из рядов третьего уровня или такое разбиение рядов на слишком большое число компонент излишнее. Если предположить, что разбиение агрегата на компоненты действительно позволяет учесть неоднородность составляющих агрегированного ряда, но оценка большого числа рядов неизбежно приводит к тому, что идиосинкритические ошибки в сумме растут, то необходимо найти компромисс между двумя этими эффектами.

Очевидно, что можно сгруппировать ряды по территориальному признаку (по странам, штатам или регионам) или по типам (для ВВП группы по отраслям и для естественного прироста отдельно ряды по рождаемости, отдельно по основным причинам смерти). Аддитивность позволяет получить ряды второго уровня просто просуммировав ряды, входящие в одну группу.

Альтернативным способом получения групп является кластеризация. Для этого использовался алгоритм иерархической кластеризации, реализованный в пакете `dwtclust`. Для получения кластеров была выбрана евклидова метрика:

$$d(y_{i,t}, y_{j,t}) = \sqrt{\sum_{t=1}^T (y_{i,t} - y_{j,t})^2}, \quad (2.6)$$

где $y_{i,t}, y_{j,t}$ – ряды нижнего уровня. Предварительно все ряды нормировались:

$$z_{ij,t} = \frac{y_{ij,t} - \bar{y}_{ij}}{\sigma(y_{ij})}, \quad (2.7)$$

где $z_{ij,t}$ – нормализованный ряд $y_{ij,t}$. Оптимальное число кластеров выбиралось так, чтобы максимизировать значение индекса силуэта SI для всей кластерной структуры:

$$SI = \frac{1}{N} \sum_{i=1}^N S_{y_i} = \frac{1}{N} \sum_{i=1}^N \frac{b_{pi} - a_{pi}}{\max(a_{pi}, b_{pi})}, \quad (2.8)$$

где N – число кластеров, S_{y_i} – силуэт каждого элемента, a_{pi} – среднее расстояние от ряда y_{pj} до других объектов из кластера y_p , b_{pi} – среднее расстояние от y_{pj} до объектов из других кластеров y_{qj} , где $q \neq p$. Для всех трех наборов данных оптимальное число кластеров по этой метрике – 25. Визуализацию рядов попавших в один кластер для квартальных данных представлены на рисунке, для сезонно сглаженных – на рисунке (2.1), для месячных – на рисунке (2.2), для всех этих графиков фиолетовые цвета означают значения ряда меньше среднего, зеленые – выше (2.3).

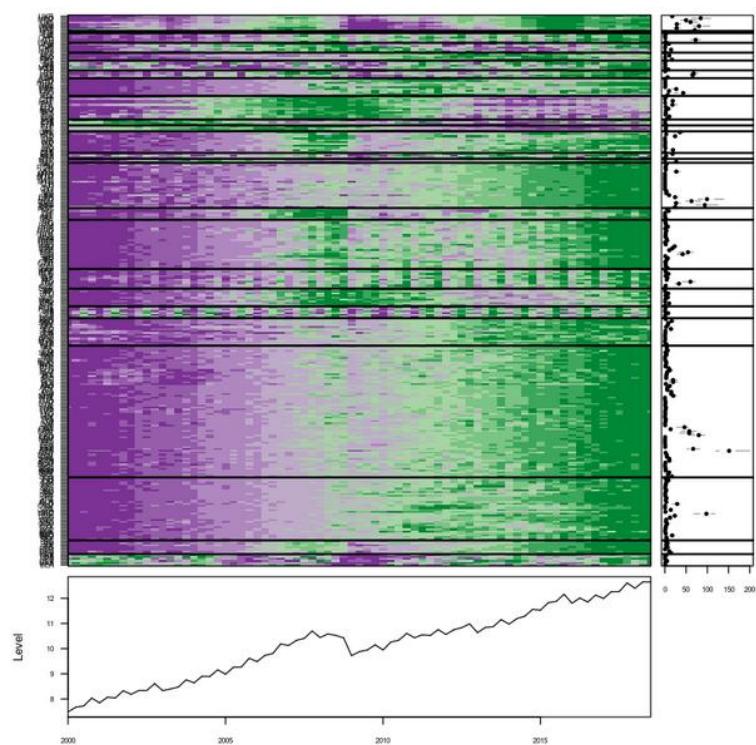


Рисунок 2.1 — Временные ряды по ВВП ЕС сгруппированные по евклидовой метрике

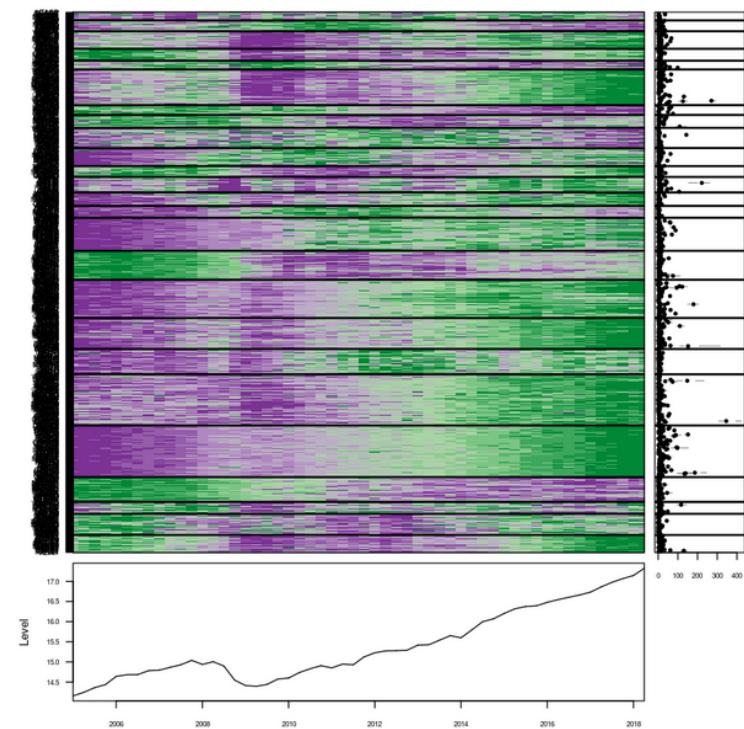


Рисунок 2.2 — Временные ряды по ВВП США сгруппированные по евклидовой метрике

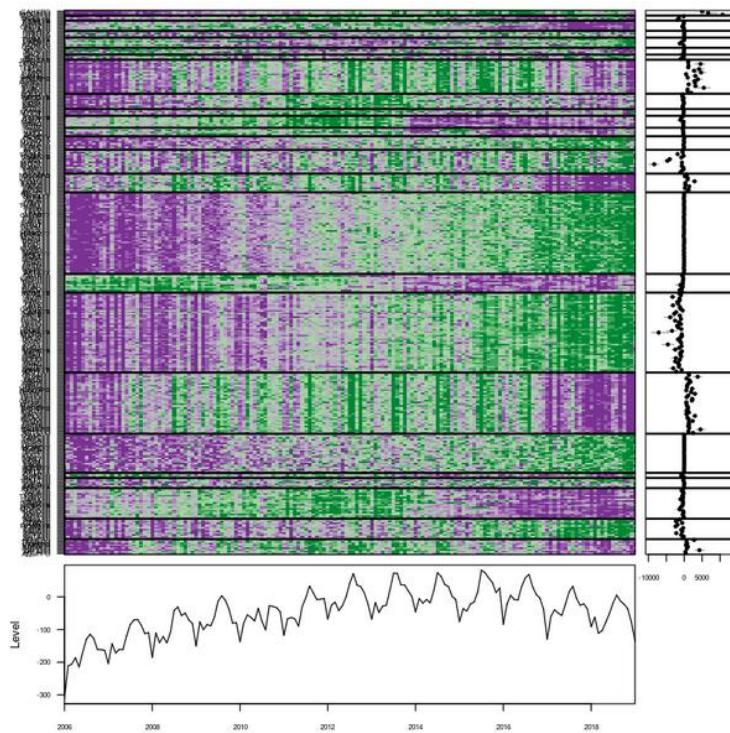


Рисунок 2.3 — Временные ряды по естественному приросту РФ сгруппированные по евклидовой метрике

При использовании процедуры кросс-валидации необходимо было на каждой итерации получать свою группировку на кластеры, но с целью проверки для экономии времени проведем кластеризацию на рядах полной длины.

2.4 Сравнение иерархических моделей

Для сравнения моделей используется следующая процедура¹²:

- разбиение рядов на подвыборки для выполнения кросс-валидации со скользящим окном с шагом в один год: для квартальных данных число подвыборок равно 6, для сезонно сглаженных – 6, для месячных – 5;
- модели описанные в таблице (2.1) используются для получения прогноза на 2 года вперед для каждого ряда, каждого уровня (отдельно с помощью пакета `hts` оценивается трехуровневая модель, отдельно три двухуровневые: сгруппированные по регионам, по классам или по кластерам);
- на каждой итерации кросс-валидации считается RMSFE для агрегированного ряда, RMSFE для невзвешенной суммы всех прогнозов нижнего ряда и RMSFE для скорректированной по OLS суммы всех прогнозов;
- для каждого набора данных RMSFE усредняется по всем подвыборкам и считается

¹²Все коды необходимые для проведения описанной процедуры выложены сайте GitHub. URL: https://github.com/xenakas/disaggregated_ts

- процентное изменение RMSFE для невзвешенной суммы и скорректированной по OLS суммы всех прогнозов по сравнению с RMSFE для агрегированного ряда;
- полученные значения для трехуровневой и двухуровневых моделей сортируются по RMSFE для агрегированного ряда.

По результатам полученным с помощью описанной процедуры формируются таблицы на которых можно увидеть, как изменился прогноз агрегированного ряда при использовании моделей, учитывающих многоуровневую структуру данных (Приложение (В)). Для наглядности при относительном уменьшении RMSFE ячейка таблицы окрашивается в зеленый, при относительном увеличении – в красный. Цвета также отличаются по интенсивности, чем ярче, тем больше отклонение от RMSFE полученного по прогнозам для агрегированного ряда.

При анализе полученных показателей и визуального представления наборов данных (Приложение (Б)) были получены три основных вывода: эффективность иерархических моделей сильно варьируется в зависимости от структуры рядов-компонент; комбинирование прогнозов с помощью OLS-корректировки позволяет устраниТЬ сильное отклонение невзвешенной суммы прогнозов от прогноза агрегированного, появившееся из-за случайного накопления идиосинкразических ошибок; предварительная группировка рядов нижнего уровня перед прогнозированием практически во всех случаях приносит положительный результат, по сравнению с прогнозами полученными по трехуровневой модели.

Опишем подробнее полученные результаты и причины, по которым были сделаны соответствующие выводы. Для невзвешенных прогнозов результаты неоднозначны: для квартальных данных в большинстве случаев при использовании иерархических моделей наблюдается ухудшение по сравнению прогнозом агрегированного ряда (2.4), для сезонно сглаженных рядов для большинства моделей наблюдается улучшение прогнозов (2.5), а для месячных на некоторых моделях наблюдается улучшение для всех вариантов структуры (трех- и двухуровневой), на некоторых – ухудшение (2.6).

Возможно это объясняется следующими фактами. Для квартальных данных все ряды имеют примерно одинаковую структуру и ошибки прогнозов не уравновешивают друг друга, а накапливаются. Для сезонно сглаженных рядов для всех итераций кросс-валидации метрика ME значительно больше нуля, то есть на всех итерациях модель в среднем занижает прогнозы. При прогнозировании отдельных компонент с этим можно бороться, поскольку только малая часть прогнозов рядов приводит к тому, что тренд недооценивается, и эти прогнозы выравниваются большим числом прогнозов улавливающих положительный тренд. Для месячных данных причина неоднозначных результатов заключается в том, что примерно четверть рядов имеет \cap -образный тренд (ряды по рождаемости), половина положительный линейный тренд, четверть отрицательный линейный тренд (что видно на рисунке (2.7), где

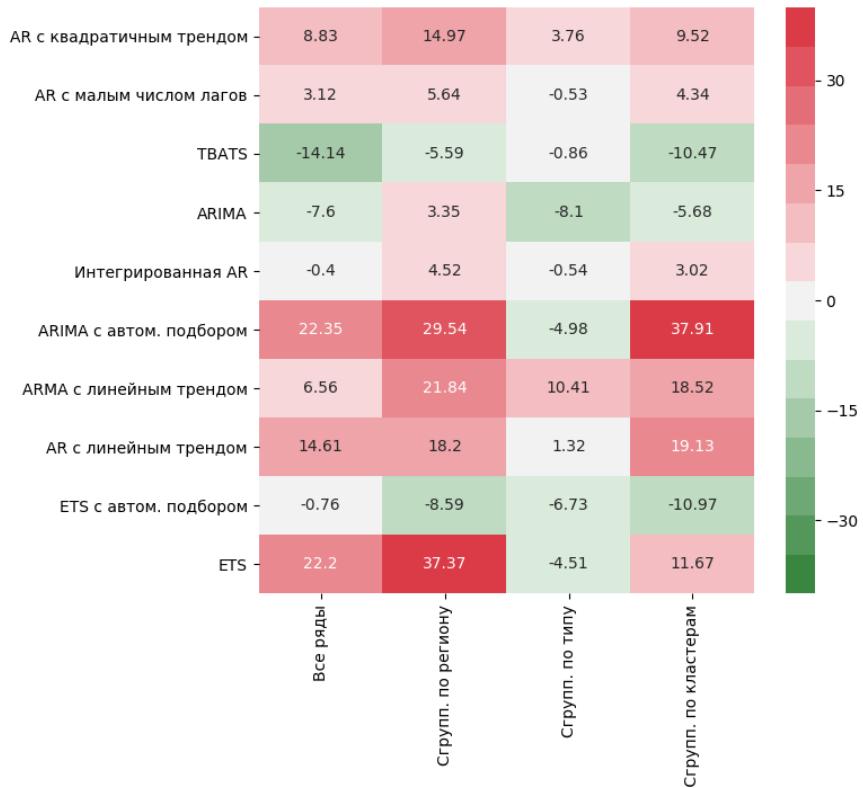


Рисунок 2.4 — Процентное изменение RMSE (невзвешенные прогнозы, квартальные ряды).

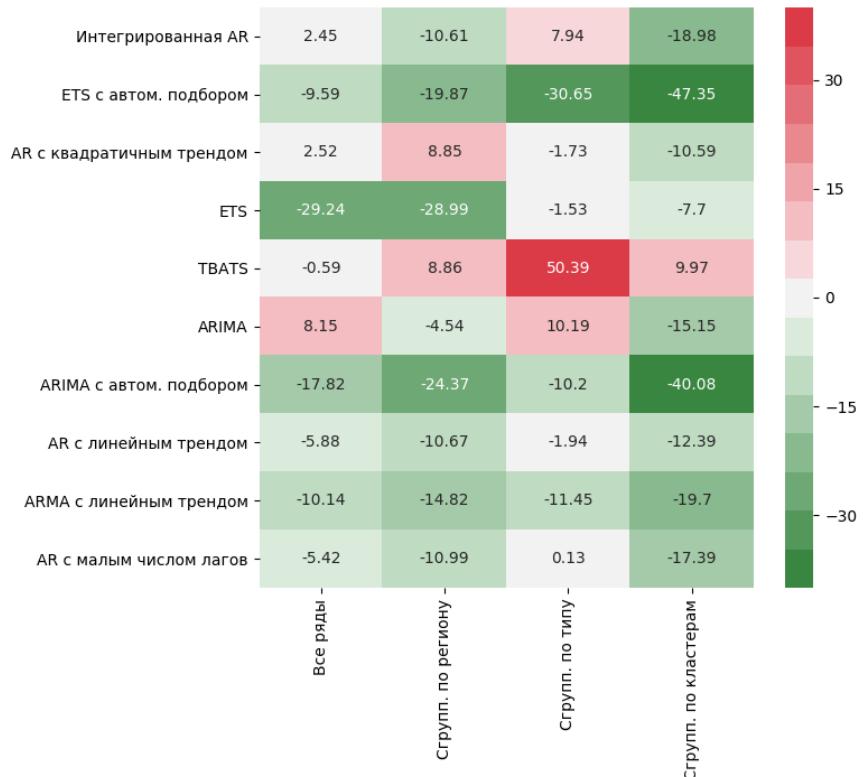


Рисунок 2.5 — Процентное изменение RMSE (невзвешенные прогнозы, сезонно сглаженные ряды).

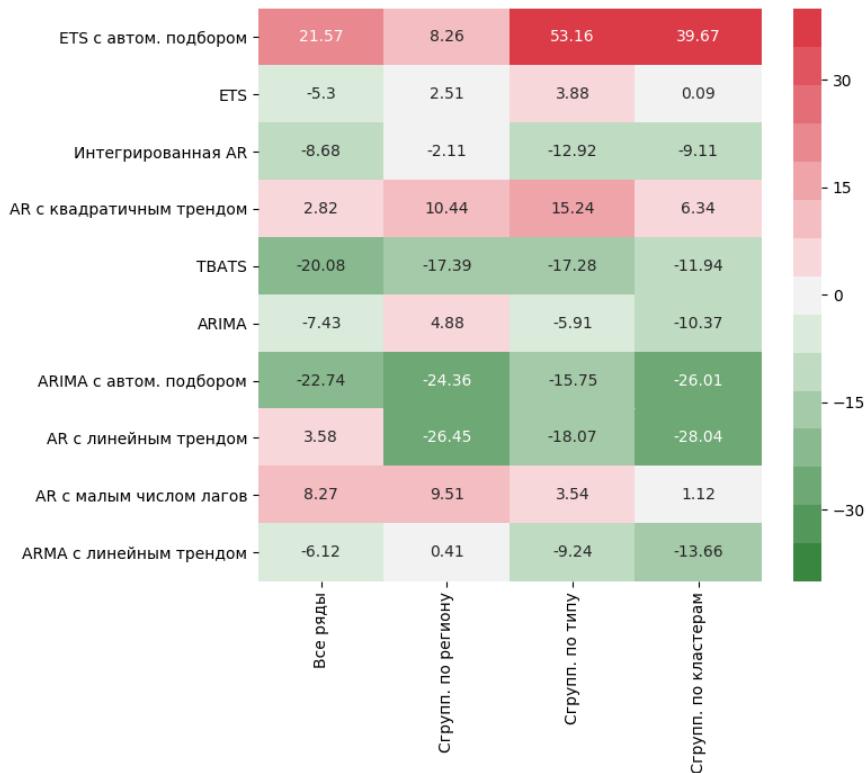


Рисунок 2.6 — Процентное изменение RMSE (невзвешенные прогнозы, месячные ряды).

ряды группируются по типу). Соответственно для некоторых моделей хорошие прогнозы имела большая по размеру группа, а для некоторых меньшая.

Прогнозы полученные с помощью OLS корректировки для квартальных и сезонно сглаженных рядов в большинстве случаев не отличаются от прогнозов полученных при прогнозировании агрегированного ряда. Доля прогнозов нижних рядов значительно отличающихся от прогнозов верхних рядов оказалась небольшой, поэтому оценки для агрегата скорректировались незначительно.

Для невзвешенных прогнозов наблюдалось резкое ухудшение для квартальных и резкое улучшение для сезонно сглаженных рядов, но при корректировке эти резкие изменения сгладились (рисунки (2.8)), (2.9))). Учитывая это можно сказать, что при небольшом числе наблюдений, недостаточном для проведения кросс-валидации, стоит использовать OLS корректировку прогнозов. Это позволит избежать резких отклонений от настоящих значений ряда, произошедших из-за случайного накопления ошибок. Накопление ошибок может привести, как к положительному эффекту, так и к отрицательному результату.

Для месячных данных при OLS корректировке в отличие от невзвешенной суммы прогнозов практически для любой модели наблюдается уменьшение RMSE по сравнению с RMSE для агрегированного ряда. Исключение составляет только модель с квадратичным трендом. Причина этому заключается в том, что только четверть рядов имеют \cap -образную форму, такую же, как у агрегированного ряда. Поэтому корректировка прогнозов агрегиро-

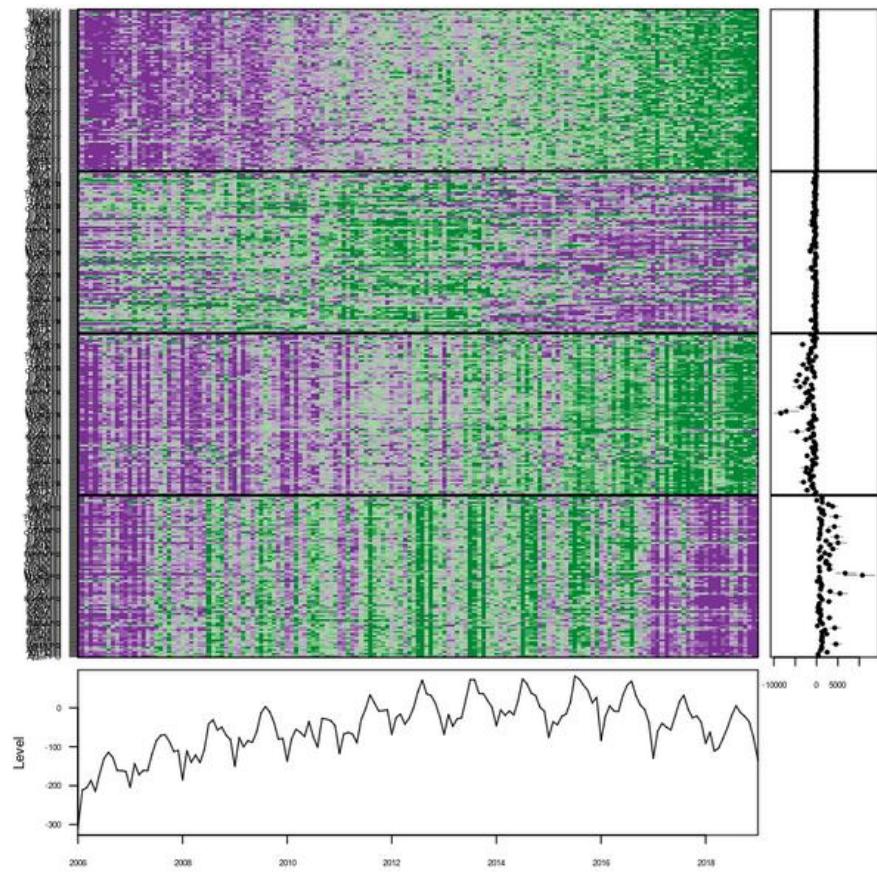


Рисунок 2.7 — Месячные ряды, сгруппированные по типу.



Рисунок 2.8 — Процентное изменение RMSE (OLS - скорректированные прогнозы, квартальные ряды).



Рисунок 2.9 — Процентное изменение RMSE (OLS - скорректированные прогнозы, сезонно сглаженные ряды).



Рисунок 2.10 — Процентное изменение RMSE (OLS - скорректированные прогнозы, месячные ряды).

ванного ряда на априори худшие прогнозы по рядам имеющим значительно отличающийся тренд приводят к тому, что прогнозы сильно корректируются в сторону ухудшения.

Если сравнивать эффективность использования различных видов группировки можно сказать, что в среднем ее использование перед получением прогнозов отдельных рядов приносит положительный результат. Причем для квартальных данных лучше всего работает группировка по отраслям (как со взвешиванием, так и без), для сезонно сглаженных для обоих способов заметное улучшение наблюдается при группировке по кластерам, для месячных данных для невзвешенных прогнозов более эффективна группировка по кластерам, а для взвешенных – группировка по типам.

В модели с комбинированием прогнозов согласно уравнению линейной регрессии (1.4), корректировка происходила с учетом трехуровневой структуры, где ряды второго уровня собирались именно по территориальному признаку. При анализе сравнительных таблиц по моделям было выявлено, что для всех наборов данных группировка по различным признакам приводит к неодинаковым результатам. По этой причине имело бы смысл проводить OLS корректировку прогнозов по суммирующей матрице, учитывающей, что на втором уровне иерархической структуре будут ряды, сгруппированные по наиболее подходящему для каждого набора данных признаку.

Во всех случаях группировка по территориальному признаку оказывалась чуть хуже других типов группировок и по отношению к трехуровневой модели без группировки улучшение происходило не во всех случаях. При группировке по этому признаку ряды не содержат в себе информацию о структуре, составляющих их компонент. Если группы по типам ведут себя похожим образом, то деления на регионы достаточно, чтобы улучшить прогнозы агрегированного ряда. Но в некоторых случаях, например, для месячных данных, ряды четко делятся на небольшое число групп примерно равного размера, потеря этой информации ведет к ухудшению прогнозов. Также деление рядов по географическому признаку может усилить эффект накапливания индивидуальных ошибок из-за неучтеною пространственной зависимости.

При группировке по типу число рядов второго уровне получается меньше, чем при группировке по территориальному признаку или по кластерам. Эти ряды не только статистически похожи друг на друга, но и имеют схожий экономический смысл. Логично предположить, что реакция таких рядов на внешние непрогнозируемые шоки с большей вероятностью будет похожей.

Группировка по кластерам в общем приносит положительные результаты, однако нельзя забывать, что за определенный отрезок времени ряды могли случайно попасть в один кластер, поскольку согласно выбранной метрике кластеризации учитывалась только близость нормированных рядов без учета каких-либо экономических факторов. По этой причине то, что в некоторых случаях прогнозы полученные после группировки по кластерам

оказывалась хуже других, можно объяснить тем, что алгоритм уловил зависимости, которых на самом деле нет.

Заключение

В этом исследовании продемонстрированы различные подходы к прогнозированию временных рядов с многоуровневой иерархической структурой, позволяющие учитывать взаимозависимости как между самими рядами, так и между прогнозами.

В результате проведенного анализа были получены три основных вывода:

- эффективность прогнозирования агрегированных рядов с помощью моделей, учитывающих многоуровневую структуру данных, сильно варьируется в зависимости от структуры наборов данных и характеристик рядов-компонент по отдельности;
- комбинирование прогнозов с помощью OLS-корректировки имеет смысл при небольшом числе наблюдений, недостаточном для проведения кросс-валидации, поскольку позволяет устраниТЬ сильное отклонение невзвешенной суммы прогнозов от прогноза агрегированного ряда по причине случайного накопления идиосинкразических ошибок;
- предварительная группировка рядов нижнего уровня перед прогнозированием практически во всех случаях дает положительный результат, по сравнению с прогнозами полученными по трехуровневой модели.

Список литературы

1. Aghabozorgi S., Shirkhorshidi A. S., Wah T. Y. Time-series clustering—A decade review // *Information Systems*. — 2015. — Vol. 53. — P. 16–38.
2. Astakhova N. N., Demidova L. A., Nikulchev E. V. Forecasting method for grouped time series with the use of k-means algorithm // arXiv preprint arXiv:1509.04705. — 2015.
3. Athanasopoulos G., Ahmed R. A., Hyndman R. J. Hierarchical forecasts for Australian domestic tourism // *International Journal of Forecasting*. — 2009. — Vol. 25, № 1. — P. 146–166.
4. Clark J. S. Uncertainty and variability in demography and population growth: a hierarchical approach // *Ecology*. — 2003. — Vol. 84, № 6. — P. 1370–1381.
5. Cobb M. Forecasting Economic Aggregates Using Dynamic Component Grouping // University Library of Munich. — 2017. — Vol. 8. — P. 3–39.
6. Cobb M. Joint forecast combination of macroeconomic aggregates and their components // University Library of Munich. — 2017. — Vol. 3. — P. 2–44.
7. Diebold F. X., Pauly P. The use of prior information in forecast combination // *International Journal of Forecasting*. — 1990. — Vol. 6, № 4. — P. 503–508.
8. Duncan G. T., Gorr W. L., Szczypula J. Forecasting analogous time series // *Principles of forecasting*. — 2001. — Vol. 3. — P. 195–213.
9. Fox D. R. Concepts and Methods of the U.S. National Income and Product Accounts. — Bureau of Economic Analysis (BEA), 2017. — P. 27–39.
10. Gelman A. Multilevel (hierarchical) modeling: what it can and cannot do // *Technometrics*. — 2006. — Vol. 48, № 3. — P. 432–435.
11. Green K. C., Armstrong J. S. Structured analogies for forecasting // *International Journal of Forecasting*. — 2007. — Vol. 23, № 3. — P. 365–376.
12. Hyndman R. J., Athanasopoulos G. *Forecasting: principles and practice*. — OTexts, 2018. — P. 456–523.
13. Hyndman R. J., Lee A. J., Wang E. Fast computation of reconciled forecasts for hierarchical and grouped time series // *Computational Statistics & Data Analysis*. — 2016. — Vol. 97. — P. 16–32.
14. Katz A. J. An Overview of BEA's Source Data and Estimating Methods for Quarterly GDP. — 10th OECD-NBS Workshop on National Accounts, 2006. — P. 145–189.
15. Makridakis S., Hibon M. The M3-Competition: results, conclusions and implications // *International journal of forecasting*. — 2000. — Vol. 16, № 4. — P. 451–476.
16. McNeish D., Wentzel K. R. Accommodating small sample sizes in three-level models when the third level is incidental // *Multivariate behavioral research*. — 2017. — Vol. 52, № 2. — P. 200–215.
17. Moyer B. C., Thompson S. Gross Domestic Product by State Estimation Methodology. — 2017. — P. 34–49.
18. Optimal combination forecasts for hierarchical time series / R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, H. L. Shang // *Computational Statistics & Data Analysis*. — 2011. — Vol. 55, № 9. — P. 2579–2589.

19. Shang H. L., Hyndman R. J. Grouped functional time series forecasting: an application to age-specific mortality rates // *Journal of Computational and Graphical Statistics*. — 2017. — Vol. 26, № 2. — P. 330–343.
20. Shang H. L., Smith P. W. Grouped time-series forecasting with an application to regional infant mortality counts // *Social Statistics & Demography*. — 2013. — Vol. 40. — P. 5–20.
21. Stegmueller D. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches // *American Journal of Political Science*. — 2013. — Vol. 57, № 3. — P. 748–761.
22. Tobias J. L. Forecasting output growth rates and median output growth rates: A hierarchical Bayesian approach // *Journal of Forecasting*. — 2001. — Vol. 20, № 5. — P. 297–314.
23. Weiss C. Essays in Hierarchical Time Series Forecasting and Forecast Combination. — 2018.
24. Zellner A., Hong C. Forecasting international growth rates using Bayesian shrinkage and other procedures // *Journal of Econometrics*. — 1989. — Vol. 40, № 1. — P. 183–202.

Приложение А Коды для выгрузки данных

A.1 Код для выгрузки данных с Eurostat

```

library(eurostat)
library(lubridate)

a10 <- get_eurostat(id="namq_10_a10", time_format="date")
a10 <- subset(a10, a10$s_adj=='NSA') # Unadjusted data
a10 <- subset(a10, a10$na_item=='B1G') #Value added, gross
a10 <- subset(a10, a10$unit=='CP_MEUR') #Current prices, million euro

a10$unit <- NULL
a10$s_adj <- NULL
a10$na_item <- NULL

gdp <- get_eurostat(id="namq_10_gdp", time_format="date")
gdp <- subset(gdp, gdp$s_adj=='NSA') # Unadjusted data
gdp <- subset(gdp, gdp$na_item=='B1G') #Value added, gross
gdp <- subset(gdp, gdp$unit=='CP_MEUR') #Current prices, million euro

gdp$unit <- NULL
gdp$s_adj <- NULL
gdp$na_item <- NULL

write.csv(a10, file = "a10_unstacked.csv")
write.csv(gdp, file = "eu_gdp_unstacked.csv")

```

A.2 Код для выгрузки данных с FRED

```

library(fredr)
library(purrr)
fredr_set_key("bd4355a0bcb45a05a7280e4d99a4d146")
readRDS("states.rds")
readRDS("industries.rds")

dict = c()
for (state in states){
    data = fredr_series_search_text(search_text =
        paste("Real Gross Domestic Product by Industry", state ),
        filter_variable = "frequency", filter_value = "Quarterly")
    data = subset(data, units == "Millions of Chained 2012 Dollars")
    dict = rbind(dict, data.frame(data$id, data$title)  )
}

id_list = as.character(dict$data.id)
gdp_us = c()
for (id in id_list){
    gdp_us = rbind(gdp_us, fredr(series_id = id )  )
}

data = fredr_series_search_text(search_text = paste("Real Gross
Domestic Product") , filter_variable = "frequency",
filter_value = "Quarterly")
data = subset(data, units == "Millions of Chained 2012 Dollars")
data = fredr_series_search_text(search_text = paste("Real Gross Domestic
Product") , filter_variable = "frequency", filter_value = "Quarterly")
data = subset(data, units == "Billions of Chained 2012 Dollars")
data = subset(data, seasonal_adjustment == "Seasonally Adjusted Annual Rate")

gdp_us = fredr(series_id = 'GDPC1' )
gdp_us$series_id <- NULL

write.csv(data, file = "naus_labelled.csv")
write.csv(gdp_us, file = "new_gdp_total.csv")

```

Приложение Б Визуализация временных рядов с трехуровневой структурой

Рисунок Б.1 — Временные ряды полученные путем агрегирования рядов третьего уровня

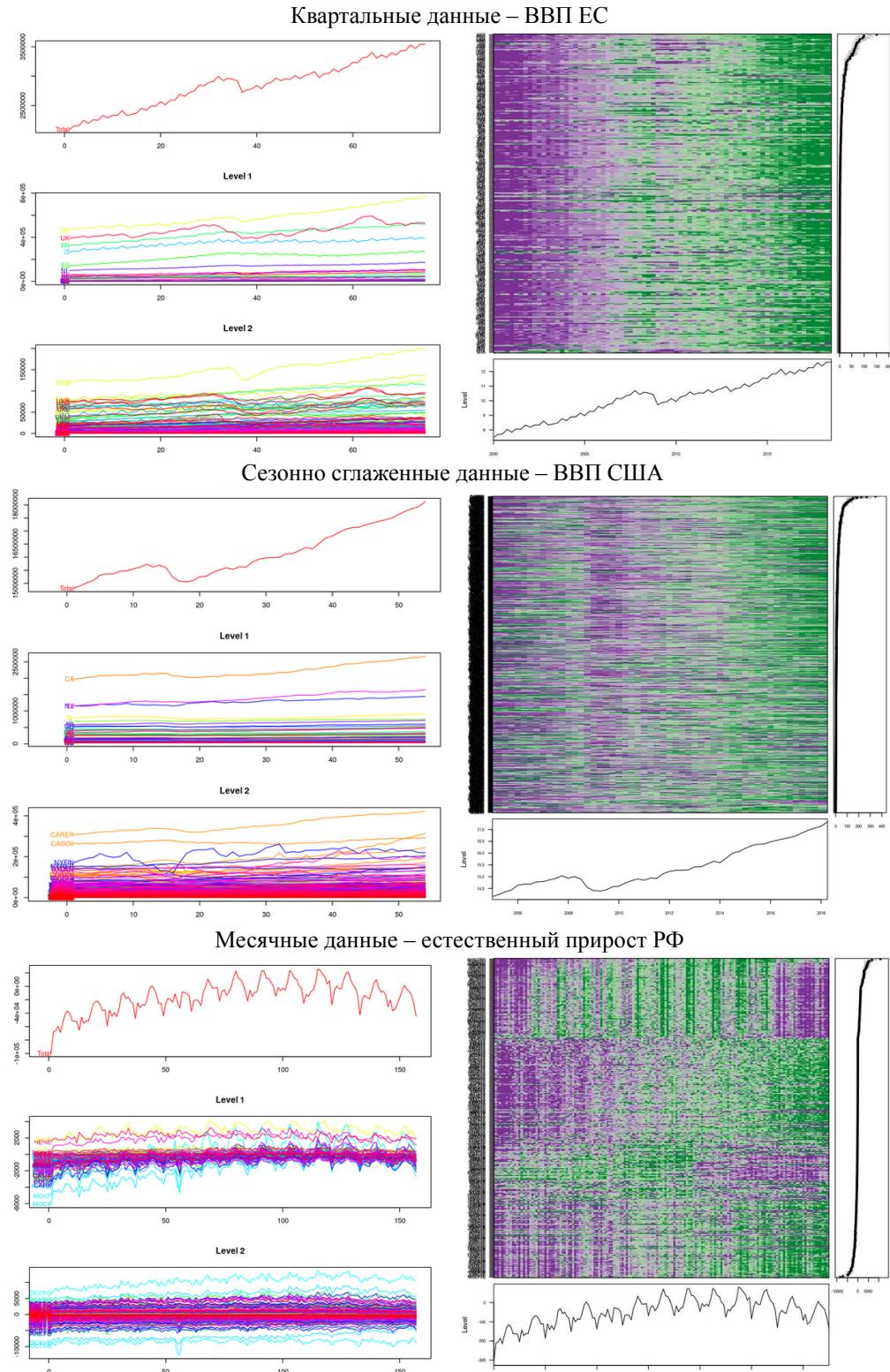
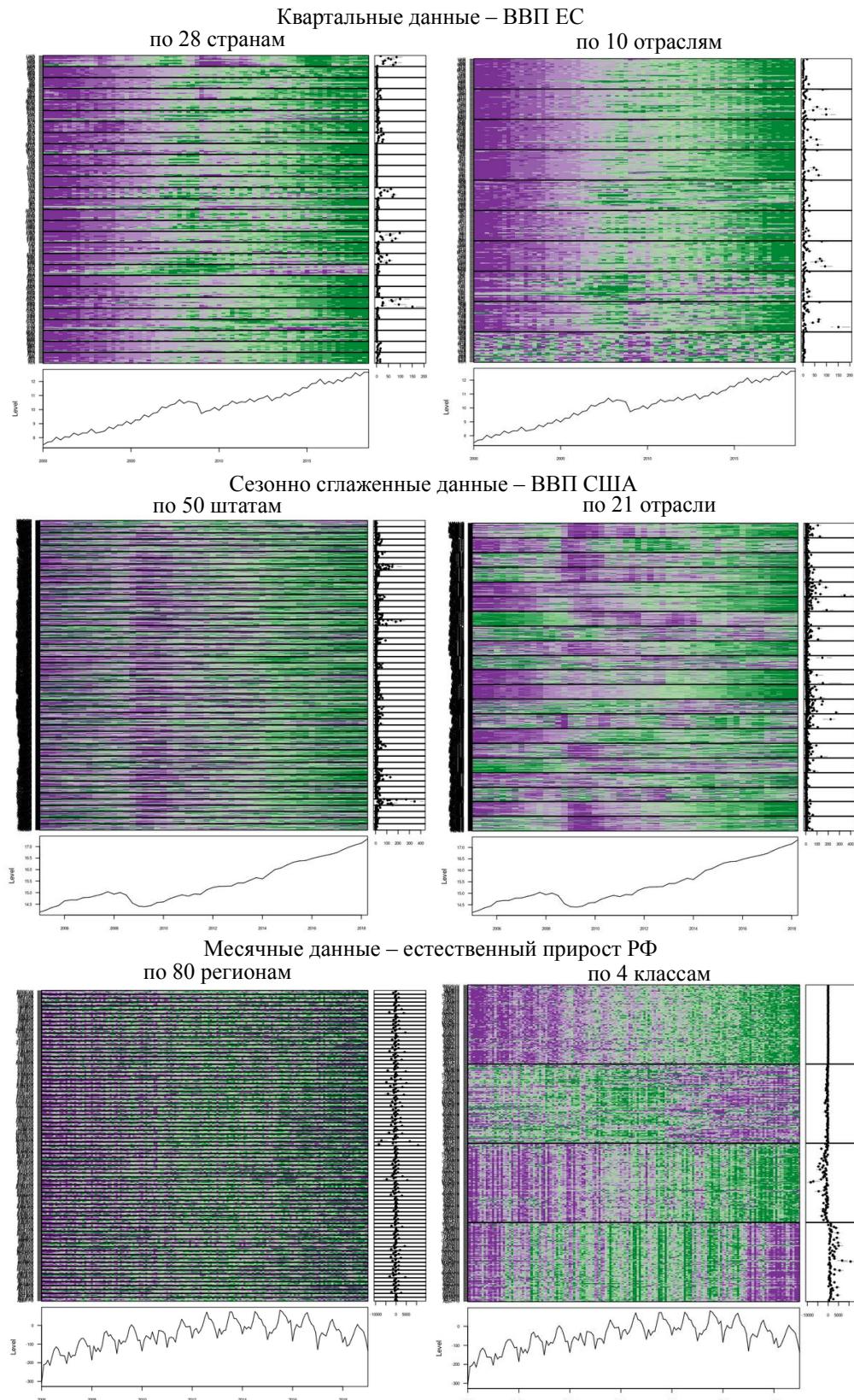
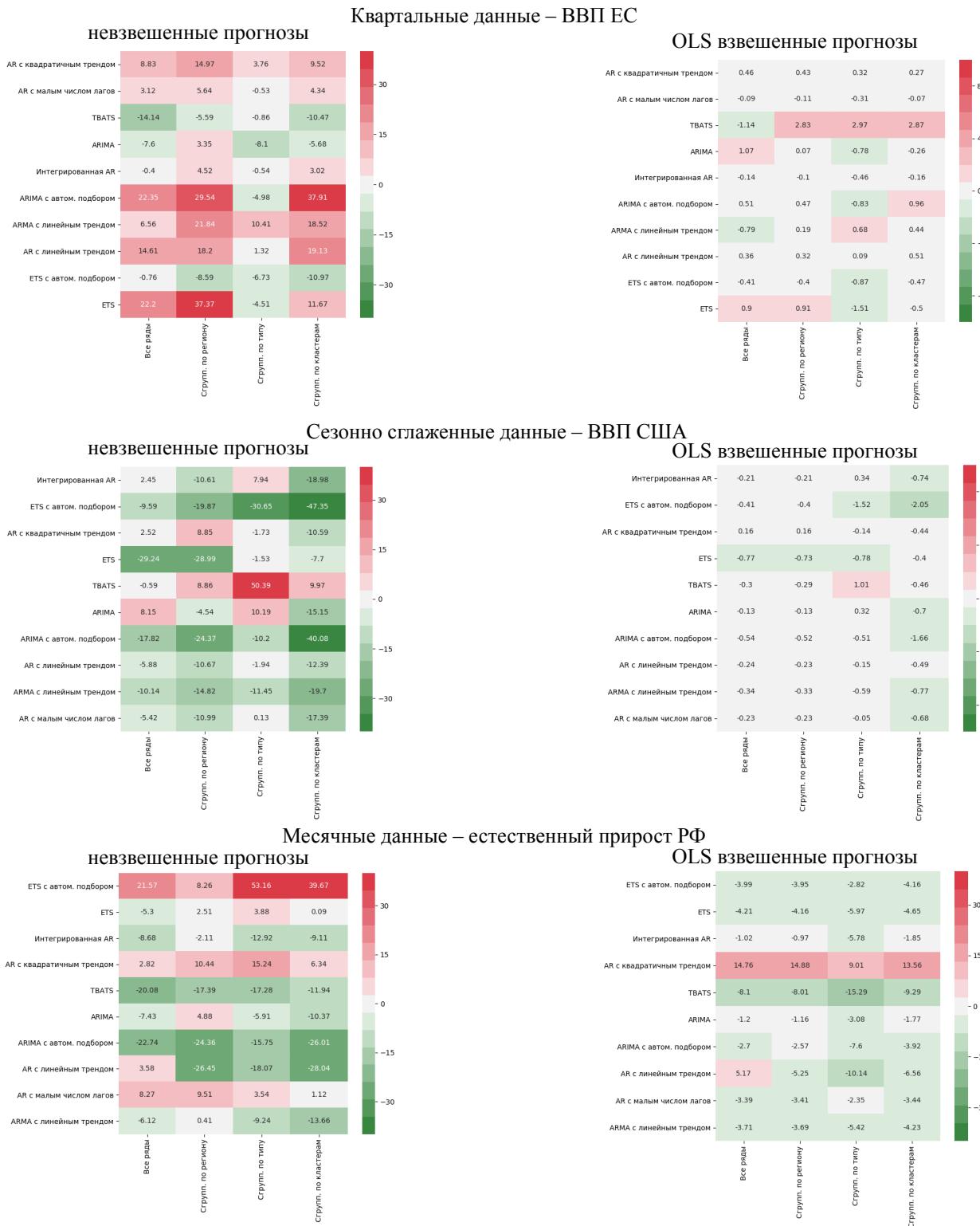


Рисунок Б.2 — Временные ряды сгруппированные по территориальному признаку и по типу



Приложение В Сравнение иерархических моделей

Рисунок В.1 — Таблицы, указывающие на процентное изменение RMSE иерархических моделей по сравнению с RMSE моделей прогнозирования агрегированного ряда



Выпускная квалификационная работа выполнена мной совершенно самостоятельно.
Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

Объем работы ____ листа(ов).

Объем приложений ____ листа(ов).

« ____ » _____ 20 ____ г.

(подпись) / Касьянова Ксения Алексеевна /