



РАНХиГС

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

Э М И Т
И Н С Т И Т У Т
ЭКОНОМИКИ, МАТЕМАТИКИ
И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА И
ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

ПРОГНОЗИРОВАНИЕ ИЕРАРХИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ

Касьянова Ксения

ЭО-15-01

Научный руководитель: Демешев Борис Борисович

Основные гипотезы

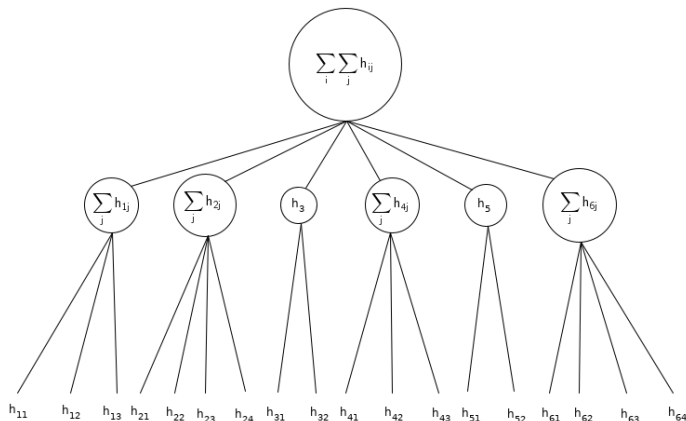


Рис.: Иерархическая структура временных рядов, необходимых для анализа

Анализ предметной отрасли

Авторы, год	Название работы	Источник	Результат
M. Cobb (2017)	Joint Forecast Combination of Macroeconomic Aggregates and Their Components	MPRA Paper from University Library of Munich	Использование оптимальной комбинации прогнозов дает более точные прогнозы агрегированного ряда, по сравнению с суммой прогнозов каждого ряда по отдельности.
R. Hyndman et al. (2010)	Optimal combination forecasts for hierarchical time series	Computational Statistics & Data Analysis	Корректировка прогнозов с помощью OLS позволяет сохранить иерархическую структуру рядов
Justin L. Tobias (2001)	Forecasting output growth rates and median output growth rates: a hierarchical Bayesian approach	Journal of Forecasting	Прогнозы можно улучшить за счет стягивания оценок коэффициентов для нижних рядов к групповым оценкам.

Цели и задачи

Актуальность:

- ▶ необходимость прогнозирования иерархических рядов в микроэкономике, макроэкономике, страховании, демографии.

Цель:

- ▶ сравнение моделей, учитывающих иерархическую структуру данных.

Задачи:

- ▶ сбор данных;
- ▶ выбор моделей;
- ▶ прогнозирование рядов второго и третьего уровня;
- ▶ сравнение суммы и оптимальной комбинации прогнозов.

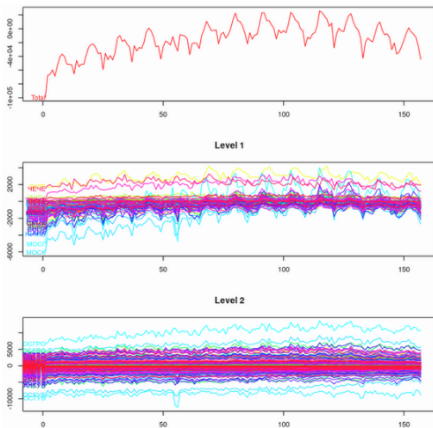
Сбор данных с иерархической структурой

Агрегированный ряд	Ряды второго уровня			Число рядов третьего уровня
	по регионам	по типам	по кластерам	
ВВП ЕС	28	10	25	280
ВВП США	50	21	25	1050
ЕП РФ	80	4	25	320

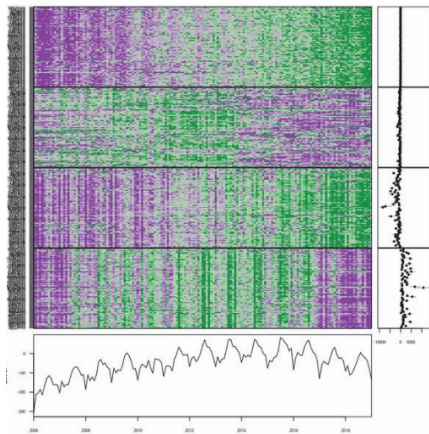
Агрегированный ряд	Сезонность	Число наблюдений	Кросс-валидация	
			число подвыборок	ширина окна
ВВП ЕС	4	75	6	48
ВВП США	4	54	6	28
ЕП РФ	12	157	5	84

Сравнение суммы и оптимальной комбинации

Визуализация месячных временных рядов



Три уровня агрегации



Группировка по типам

Таблица: Параметры моделей

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

(p,d,q) – параметры для ARIMA модели, $(P,D,Q)_m$ – параметры для SARIMA модели с периодичностью m , A - аддитивный, Ad - аддитивный демпфированный, M - мультипликативный.

Сравнение суммы и оптимальной комбинации

Оптимальная комбинация прогнозов

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

Сравнение суммы и оптимальной комбинации

Оптимальная комбинация прогнозов

$$y_t = Sb_t,$$

где y_t – вектор всех наблюдений на всех уровнях иерархии в момент времени t , S – суммирующая матрица b_t – вектор всех наблюдений на самом нижнем уровне иерархии в момент времени t

Невзвешенная сумма:

$$\tilde{y}_h = S\hat{y}_{K,h},$$

OLS-корректировка:

$$\hat{y}_h = S\beta_h + e_h,$$

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

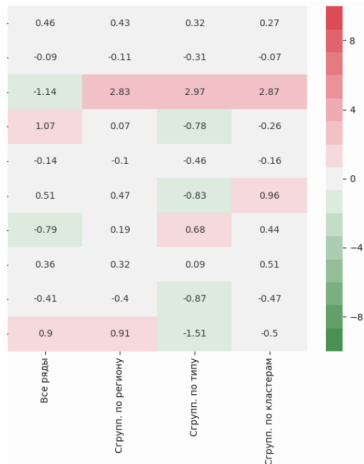
где \tilde{y}_h – собранные с помощью суммирования прогнозы рядов уровней $1...K-1$ и базовые прогнозы $\hat{y}_{K,h}$.

Сравнение суммы и оптимальной комбинации

Процентное изменение RMSE для квартальных рядов



Невзвешенная сумма



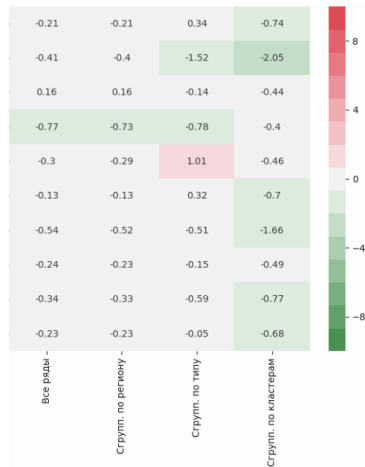
OLS-корректировка

Сравнение суммы и оптимальной комбинации

Процентное изменение RMSE для квартальных сезонно-сглаженных рядов



Невзвешенная сумма



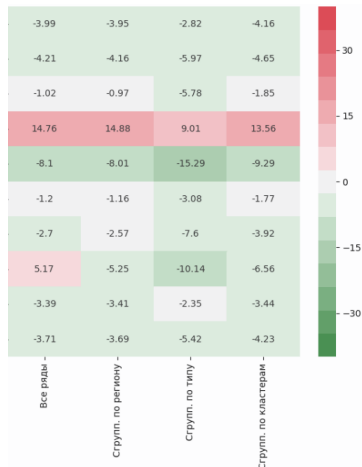
OLS-корректировка

Сравнение суммы и оптимальной комбинации

Процентное изменение RMSE для месячных рядов



Невзвешенная сумма



OLS-корректировка

Результат:

- ▶ эффективность прогнозирования сильно варьируется в зависимости от характеристик наборов данных и их рядов-компонент;
- ▶ корректировка прогнозов с помощью OLS позволяет избавиться от случайного накопления идиосинкразических ошибок;
- ▶ группировка рядов улучшает прогнозы по сравнению с прогнозами полученными по трехуровневой модели.

Спасибо за внимание!

ПРОГНОЗИРОВАНИЕ ИЕРАРХИЧЕСКИХ
ВРЕМЕННЫХ РЯДОВ

Касьянова Ксения

ЭО-15-01

Актуальность:

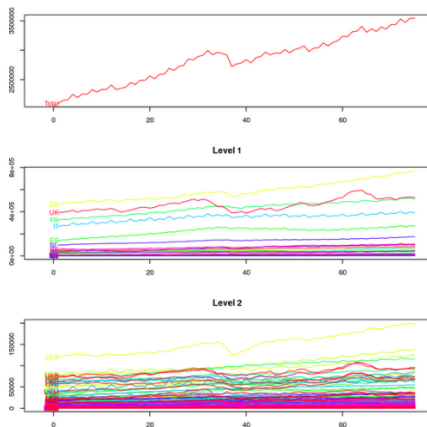
- ▶ точный прогноз экономических показателей – один из ключевых факторов принятия эффективных решений;
- ▶ необходимость прогнозирования данных с иерархической структурой в микроэкономике, макроэкономике, страховании, демографии;
- ▶ применение иерархических моделей на трех наборах данных с использованием кросс-валидации позволит получить более устойчивые выводы.

Гипотезы:

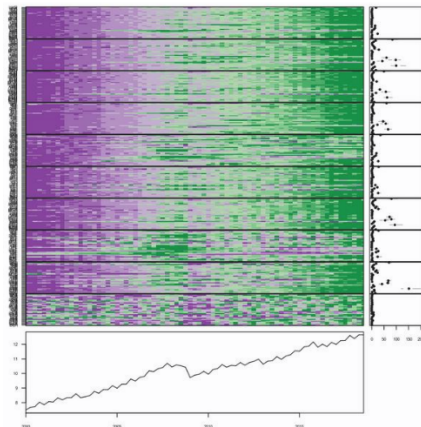
- ▶ при группировке рядов, которые ведут себя одинаково, идиосинкразические ошибки внутри групп будут компенсировать друг друга, в то время общая динамика будет сохраняться;
- ▶ подбор весов с помощью регрессии позволяет учесть более точные прогнозы с большим весом.

Сравнение суммы и оптимальной комбинации

Визуализация квартальных временных рядов



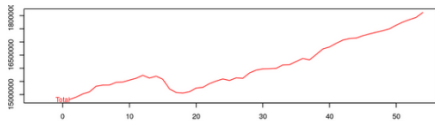
Три уровня агрегации



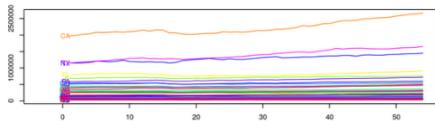
Группировка по отраслям

Сравнение суммы и оптимальной комбинации

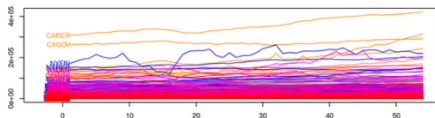
Визуализация квартальных сезонно-сглаженных временных рядов



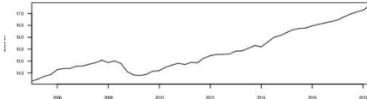
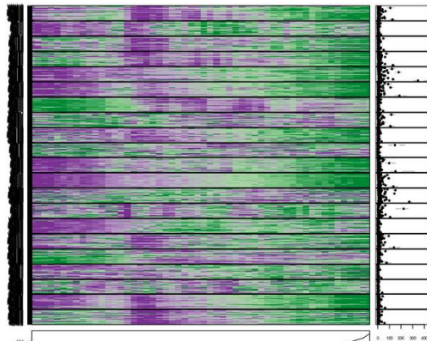
Level 1



Level 2



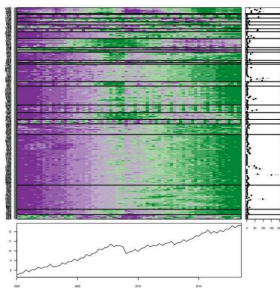
Три уровня агрегации



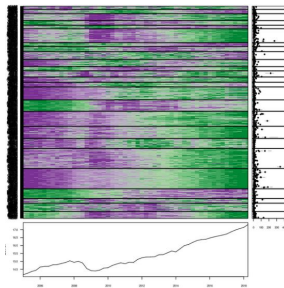
Группировка по отраслям

Сравнение прогнозов

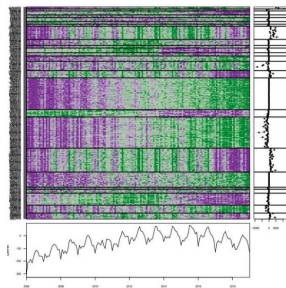
Временные ряды сгруппированные по евклидовой метрике



ВВП ЕС



ВВП США



ЕП РФ

Сравнение суммы и оптимальной комбинации

Квадратный корень из среднеквадратичной ошибки (root mean square error):

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

Евклидова метрика:

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

Индекс силуэта для кластерной структуры:

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