

Прогнозирование индексов цен с использованием больших массивов данных

Е. Астафьева, В. Петренко, М. Турунцева

16 января 2014 г.



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

Методика

ШАГ 1. Преобразование исходных данных: преобразование временных рядов к стационарному виду; исключение (корректировка) выбросов; нормировка полученных временных рядов.

ШАГ 2. Снижение размерности исходного массива данных (МГК).

ШАГ 3. Оценка моделей для прогнозирования.

ШАГ 4. Построение прогнозов.

Данные

- Исходный массив: 113 рядов с января 2002 г. (8 групп показателей)
- ИПЦ, ИЦП ПТ, ИЦП ПТ в: добыче полезных ископаемых; обрабатывающих производствах; производстве и распределении электроэнергии, газа и воды; производстве пищевых продуктов; текстильном и швейном производстве; обработке древесины и производство изделий из дерева; целлюлозно-бумажном производстве; производстве кокса, нефтепродуктов; химическом производстве; металлургическом производстве и производстве готовых металлических изделий; производстве машин и оборудования; производстве транспортных средств и оборудования
- Сравнение с альтернативными методами: прогнозы ИЭП (прогнозы в режиме «реального времени»), наивные, наивные сезонные, скользящее среднее

Блок 1 – зависимость от горизонта прогнозирования

- Индекс потребительских цен
- Прогнозы на январь-декабрь 2012 г. (12 месяцев), апрель-декабрь 2012 г. (9 месяцев), июль-декабрь 2012 г. (6 месяцев) и октябрь-декабрь 2012 г. (3 месяца)

Блок 1 – зависимость от горизонта прогнозирования:

Число факторов	Лаги факторов	Лаги объясняемой переменной	Индекс потребительских цен			
			12 мес. (01.12 – 12.12)	9 мес. (04.12 – 12.12)	6 мес. (07.12 – 12.12)	3 мес. (10.12 – 12.12)
прогноз по ARIMA-моделям			0,36%	0,22%	0,27%	0,11%
1	12	1, 12				
1	12	1, 6, 12				
1	13	1	0,31%			
1	13	1, 6, 12				
1	12, 13	1, 6, 12				
1	12, 14	12	0,35%			
1	12, 14	1, 6, 12				
1	12, 15	1, 6, 12				
1	12, 13, 15	1, 6, 12				
2	12, 13	12				
2	12, 13	1, 6, 12				0,06%
2	12, 13, 14	12				
2	12, 13, 14	1, 12				0,05%
2	12, 13, 14	1, 6, 12				0,05%
3	13	12	0,34%			
3	12, 13	1, 6, 12				
3	12, 13, 14	1, 12				
3	12, 13, 14	1, 6, 12				0,02%
4	13	12				
4	12, 13, 14	1, 6, 12		0,23%		
4	12, 13, 18	1		0,23%	0,20%	
4	12, 13, 18	1, 12		0,24%	0,19%	
4	12, 13, 18	1, 6, 12			0,19%	



Блок 1 – зависимость от горизонта прогнозирования: выводы

- качество прогнозирования по всем моделям ухудшается с увеличением горизонта прогнозирования;
- практически всегда факторные прогнозы оказываются лучше любого из альтернативных прогнозов;
- разрыв между качеством факторных прогнозов и ARIMA–прогнозов уменьшается с увеличением горизонта прогнозирования;
- не было найдено факторной модели, которая давала бы лучший прогноз на всех горизонтах прогнозирования.

Блок 2 – сравнительный анализ качества

- 14 рядов
- Прогнозы на июль-декабрь 2012 г.

Блок 2 – сравнительный анализ качества: ВЫВОДЫ

- В половине случаев (7 из 14) лучшей для целей прогнозирования оказалась факторная модель.
- Для трех из этих семи моделей число включенных в лучшую модель факторов совпало с оптимальным (всего совпало 5 из 14).
- Прогнозы, построенные на основе скользящего среднего, были лучшими для четырех показателей из 14.
- Наивный сезонный прогноз оказался лучшим для двух рядов, наивный – для одного.



Блок 3 – устойчивость прогнозирования отдельных месяцев

- Индекс потребительских цен
- Январь 2009 – июль 2012 г.



Блок 3 – устойчивость прогнозирования отдельных месяцев: выводы

- несмотря на то что для большинства месяцев лучшие прогнозы были получены при помощи факторных моделей, нам не удалось найти *одну* факторную модель, которая давала бы устойчиво лучший результат для большинства месяцев;
- на всем массиве данных ARIMA–прогноз оказывается лучше, чем факторный прогноз, построенный по модели, являющейся лучшей с точки зрения качества прогнозирования на всем массиве.

Общие выводы

- Лучшие прогнозы, построенные с использованием больших массивов данных, как правило оказываются лучше альтернативных методов прогнозирования.
- Тем не менее, не удалось найти одну модель (для каждого ряда), которая была бы устойчиво лучшей для всех моментов прогнозирования и всех горизонтов прогнозирования.
- Оптимальное число факторов в модели по критерию Баи-Нг не всегда приводит в лучшему прогнозу (а чаще всего этого не происходит).

Прогноз на 2013 г.

	Июнь 2013	Июль 2013	Август 2013	Сентябрь 2013	Октябрь 2013	Ноябрь 2013	Средний
ARIMA	8,2%	8,3%	5,4%	7,7%	6,9%	6,0%	7,1%
SM	5,9%	6,0%	5,8%	5,9%	6,1%	6,3%	6,0%
Средний	7,1%	7,2%	5,6%	6,8%	6,5%	6,2%	6,54%
Факт	6,5%						



Прогноз на 2014 г.

5,6-6,1%

Спасибо за внимание!



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