

ПРОГНОЗИРОВАНИЕ ЦЕН НА ТОВАРЫ И УСЛУГИ ПРЕДПРИЯТИЙ КЫРГЫЗСКОЙ РЕСПУБЛИКИ

Москаленко А.А., магистрант КГТУ им. И. Раззакова, Кыргызстан, 720044, г. Бишкек, пр. Ч.Айтматова 66, e-mail: moskalenkoalim@mail.ru

Аннотация. В данной статье описывается проведение прогнозирования временных рядов цен на различные товары и услуги в Кыргызской Республике. Различные методы прогнозирования сравниваются между собой по критерию точности полученных предсказаний, полученные результаты анализируются и обобщаются.

Ключевые слова: прогнозирование временных рядов, ARIMA, двойное сглаживание Брауна, двойное сглаживание Холта, средняя абсолютная скалированная ошибка

PREDINCTION OF PRICES FOR GOODS AND SERVICES OF ENTERPRISES OF THE KYRGYZ REPUBLIC

Moskalenko A. A., second-year master student of Software Engineering (SE) Department, Kyrgyzstan, 720044, Bishkek city, KSTU named after I.Razzakov, e-mail: moskalenkoalim@mail.ru

Summary. This article describes the forecasting of time series of prices for various goods and services in the Kyrgyz Republic. Different forecasting methods are compared with each other by the measure of accuracy of the predictions. At the end, the results are analyzed and generalized.

Keywords: forecasting of time series, ARIMA, Brown double smoothing, Holt double smoothing, mean absolute scaled error

При прогнозировании временных рядов выбор оптимального метода прогнозирования сильно зависит от условий, при которых временной ряд был получен – не существует универсального метода прогнозирования, который во всех случаях даёт наиболее точные предсказания. Методы прогнозирования отличаются друг от друга и по точности предсказаний, и по сложности реализации – от простейшего наивного метода прогнозирования до комбинированных методов, состоящих из нескольких компонентов. Как показала практика [5], на данный момент методы машинного обучения, такие, как нейронные сети, дают при прогнозировании временных рядов не самые точные результаты, в особенности, если исходных данных не очень много. Зато применение статистических методов, дополненных методами машинного обучения, предсказывают временные ряды с наиболее высокой точностью [6].

В данной статье описаны результаты применения нескольких базовых методов прогнозирования к исходным статистическим данным. Данные представляют из себя ежемесячную статистику с 2003 до 2020 года по уровню цен на 77 различных категорий товаров и услуг в Кыргызской Республике [7], данные предоставляются в открытый доступ Национальным статистическим комитетом. Используемые в работе методы – ARIMA, наивный прогноз, метод простого экспоненциального сглаживания, метод двойного сглаживания Брауна, метод двойного сглаживания Холта.

Оценка точности прогнозирования

Оценить результаты прогнозирования, полученные после применения некоторого метода прогнозирования, зачастую не очень просто. Помимо того, что необходимо на одних и тех же данных сравнить точность разных методов, необходимо ещё и сравнивать точность работы методов прогнозирования на различных данных, независимо от масштаба данных, количества данных, наличия среди данных нулевых значений. Многие стандартные меры, такие, как средняя абсолютная ошибка, или средняя относительная ошибка, не удовлетворяют вышеперечисленным качествам [4]. Поэтому, для оценивания качества прогнозов, в данной работе была выбрана средняя абсолютная скалированная ошибка (MASE), удовлетворяющая всем основным благоприятным признакам [4]. Ошибка MASE показывает, насколько полученный прогноз превосходит одношаговый наивный прогноз, а расчёт ошибки происходит по формуле: $MASE = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - f_i|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}$, где n – количество прогнозных значений, $y_i, i=1...n$ – фактические значения ряда, $f_i, i=1...n$ – прогнозируемые значения ряда.

В то же время, при использовании метода наивного одношагового прогнозирования, ошибка MASE теряет смысл – в таком случае можно вычислять среднюю относительную ошибку прогнозирования (MAPE), и с помощью неё сравнивать точность прогнозирования на различных данных.

Метод наивного одношагового прогнозирования

Метод наивного одношагового прогнозирования – простейший метод прогнозирования, при котором прогнозное значение равно предыдущему фактическому значению. Несмотря на свою тривиальность, метод способен давать достаточно точные прогнозы – особенно при небольших колебаниях временного ряда относительно своих абсолютных значений. При проведении исследования средняя ошибка рассчитывалась для 5 последних месяцев из временной выборки (данные с сентября 2019 по январь 2020).

В таблице 1 приведены все категории товаров и услуг, со значениями средней относительной ошибки при прогнозировании наивным одношаговым методом.

Таблица 1.

Категории товаров и услуг, проранжированные по значениям ошибки MAPE при прогнозировании одношаговым наивным методом

№	Категория	Ошибка MAPE (%)
1	арендная плата за жилье	0,00
2	тепловая энергия	0,00
3	связь	0,00

4	почтовые услуги	0,00
....
75	фрукты	3,10
76	овощи	4,37
77	водоснабжение	4,76

Как видно из таблицы 1, для некоторых категорий услуг наивный прогноз дал абсолютно точные предсказания – это связано с тем, что статистика цен на такие услуги может быть одинаковой в течение нескольких месяцев. Это правило относится в первую очередь к услугам, предоставляемым государством.

В то же время, и для всех остальных категорий товаров и услуг средняя относительная ошибка не превышает 5%, что является достаточно точным результатом.

Метод простого экспоненциального сглаживания

Метод простого экспоненциального сглаживания – ещё один метод прогнозирования, который имеет простую структуру, прогнозное значение в данном методе рассчитывается по формуле: $f_{i+1} = \alpha y_i + (1 - \alpha) f_i$, где y_i – фактические значения временного ряда, f_i – прогнозные значения, $\alpha \in (0; 1)$ – переменный параметр.

Таблица 2.

Категории товаров и услуг, проранжированные по значениям ошибки MASE при прогнозировании методом простого экспоненциального сглаживания

	Категория	Ошибка MASE	α
1	услуги	0,326426	0,5
2	транспортные услуги	0,513847	0,1
3	малые электробытовые приборы	0,719554	0,1
4	транспорт	0,804941	0,1
....	
70	мясо	1,141193	0,9
71	газеты, книги и канцелярские товары	1,155249	0,9
72	амбулаторные услуги	1,157899	0,9

При проведении прогнозирования в большинстве случаев результаты простого экспоненциального сглаживания оказались худшими, либо незначительно превосходящими наивное одношаговое прогнозирование. Значения параметра α варьировались в пределах (0;1) с шагом 0,1, после расчёта всех прогнозных значений выбиралось то значение α , которое соответствовало наилучшему прогнозу. Часть результатов, полученных при применении метода, можно увидеть в таблице 2.

Некоторые категории товаров и услуг не попали в итоговую таблицу, так как для них наивный одношаговый прогноз давал абсолютно точные результаты, и значение ошибки MASE оказывалось в таком случае бесконечным.

Метод двойного сглаживания Брауна

Метод двойного сглаживания Брауна (ДСБ) – одна из модификаций метода простого экспоненциального сглаживания. Как правило, метод даёт более точные прогнозы, по сравнению с предыдущими вариантами [2]. Формула, по которой рассчитывается прогнозное значение: $f_{i+1} = 2 p_{i+1} - q_{i+1} + \frac{\alpha}{1-\alpha} (p_{i+1} - q_{i+1})$, $p_{i+1} = \alpha y_i + (1 - \alpha) p_i$, $q_{i+1} = \alpha p_{i+1} + (1 - \alpha) q_i$, где y_i – фактические значения временного ряда, f_i – прогнозные значения, $\alpha \in (0; 1)$ – переменный параметр.

В результате проведения прогнозирования, для 70% категорий точность оказалась выше, чем для наивного прогнозирования. Часть результатов представлена в таблице 3, пример категории («все товары и услуги»), для которой метод двойного сглаживания Брауна дал наиболее точные результаты, представлен на рисунках 1, 2.

Категории товаров и услуг, проранжированные по значениям ошибки MASE при прогнозировании методом двойного сглаживания Брауна.

№	Категория	Ошибка MASE	α
1	все товары и услуги	0,372406	0,99
2	электроэнергия, газ и прочие виды топлива	0,437023	0,05
3	сахар, джем, мед, шоколад и конфеты	0,443144	0,99
4	мясо	0,449595	0,87
****	*****	****	
70	рестораны и гостиницы	1,165660	0,84
71	мужская одежда	1,168415	0,78
72	общественное питание	1,192049	0,86



Рис.1. Прогноз ДСБ
для временного промежутка 2003-2020

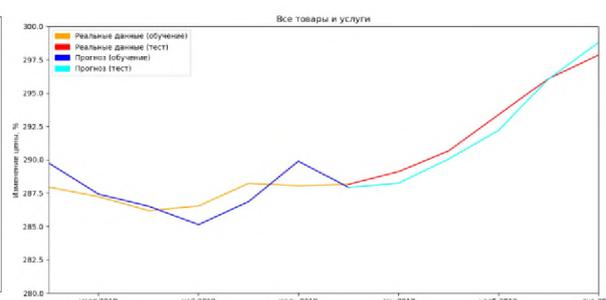


Рис.2. Прогноз ДСБ
для временного промежутка 2019-2020

Как и в случае простого экспоненциального сглаживания, оптимальный параметр α искался на интервале $(0;1)$, но в данном случае с шагом 0.01. Стоит отметить, что категории товаров и услуг, для которых методы прогнозирования оказываются самыми успешными – отличаются от метода к методу, что видно из таблиц 2 и 3.

Метод двойного сглаживания Холта

Метод двойного сглаживания Холта (ДСХ) – ещё одна из модификаций метода простого экспоненциального сглаживания [3]. Однако, формула, по которой рассчитывается прогнозное значение, имеет больше свободных параметров, чем в двойном сглаживании Брауна: $f_{i+1} = p_{i+1} + q_{i+1}$, $p_{i+1} = \alpha y_i + (1 - \alpha)(p_i + q_i)$, $q_{i+1} = \beta(p_{i+1} - p_i) + (1 - \beta)q_i$, где y_i – фактические значения временного ряда, f_i – прогнозные значения, $\alpha, \beta \in (0; 1)$ – переменные параметры.

В результате проведения прогнозирования, для 90% категорий точность оказалась выше, чем для наивного прогнозирования. Часть результатов представлена в таблице 4, пример категории («пищевые продукты и безалкогольные напитки»), для которой метод двойного сглаживания Холта дал наиболее точные результаты, представлен на рисунках 3, 4.

Таблица 4.

Категории товаров и услуг, проранжированные по значениям ошибки MASE при прогнозировании методом двойного сглаживания Холта

№	Категория	Ошибка MASE	α	β
1	пищевые продукты и безалкогольные напитки	0,165918	0,08	0,73
2	сахар, джем, мед, шоколад и конфеты	0,212444	0,14	0,88
3	жилищные услуги: вода, электроэнергия, газ и другие виды топлива	0,227260	0,04	0,52
****	*****	****		
70	личное обслуживание	0,959093	0,02	0,18
71	водоснабжение	1,000138	0,73	0,37
72	прочие услуги, не отнесенные к другим категориям	1,015287	0,73	0,37



Рис. 3. Прогноз ДСХ для временного промежутка 2003-2020



Рис. 4. Прогноз ДСХ для временного промежутка 2019-2020

Прогнозирование при помощи метода ARIMA

Одним из универсальных методов, позволяющих качественно предсказывать значения временного ряда, является метод ARIMA – модель Бокса-Дженкинса [1]. Этот метод сочетает в себе модели авторегрессии и скользящего среднего, и может быть применён к нестационарным рядам. Нестационарные ряды в этом методе приводятся к стационарным путём взятия разностей некоторого порядка. Параметры модели ARIMA (p, d, q): p – порядок авторегрессии, d – порядок взятия разности, q – порядок скользящего среднего.

В данной работе прогнозы строились при помощи модели ARIMA (2,1,2) на основании временных рядов из 200 значений (данные с января 2003 по август 2019), качество прогноза проверялось на 5 новых значениях (данные с сентября 2019 по январь 2020). Модель ARIMA (2,1,2) была выбрана как оптимальная после проверки различных моделей с другими параметрами. Прогноз строился в программе Python с использованием библиотеки statsmodels.

Таблица 5.

Категории товаров и услуг, проранжированные по значениям ошибки MASE при прогнозировании методом ARIMA

№	Категория	Ошибка MASE
1	газеты, книги и канцелярские товары	0,418677
2	женская одежда	0,591931
3	мужская одежда	0,657436
4	овощи	0,670382
....
60	женская обувь	7,207952
61	услуги по организации культурных мероприятий	7,952600
62	транспортные услуги	10,812246
63	транспорт	13,061948

В таблице 5 приведены категории товаров и услуг, со значениями ошибки MASE при прогнозировании методом ARIMA.

Как видно из таблицы 5, для некоторых категорий товаров метод ARIMA привёл к хорошим прогнозным значениям. Значение ошибки MASE, меньшее 1, означает то, что метод прогнозирования сработал лучше наивного прогноза – это случилось для 12 категорий. Для остальных категорий метод отработал хуже наивного прогноза. Отдельно стоит отметить, что для 14 непредставленных в таблице категорий метод ARIMA либо не сошёлся ни при каких параметрах, либо значение ошибки MASE оказалось равным бесконечности.

Работа метода ARIMA для категории «газеты, книги и канцелярские товары» представлена на рисунках 5, 6.

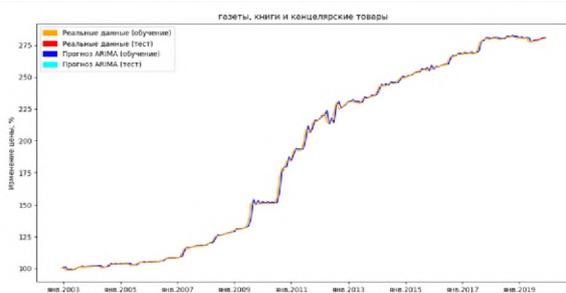


Рис.5. Прогноз ARIMA для временного промежутка 2003-2020

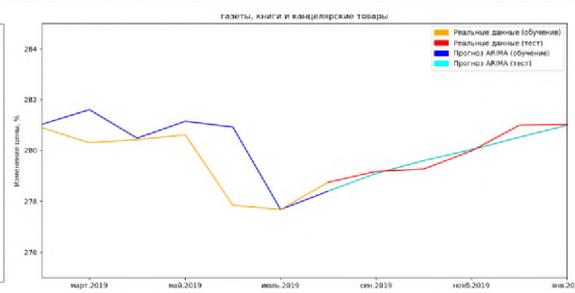


Рис.6. Прогноз ARIMA для временного промежутка 2019-2020

Сравнение работы различных методов

В результате построения прогнозов для 77 категорий товаров и услуг при помощи 5 различных методов, для каждой категории был обнаружен метод, показывающий наилучший результат при краткосрочном прогнозировании. Результаты прогнозирования для 5 последних месяцев (данные с сентября 2019 по январь 2020) представлены в таблице 6.

Таблица 6.

Категории товаров и услуг, проранжированные по значениям ошибки MASE для различных методов прогнозирования за период 2019-2020

№	Категория	мин. ошибка	Метод
1	пищевые продукты и безалкогольные напитки	0,165918	ДСХ
2	сахар, джем, мед, шоколад и конфеты	0,212444	ДСХ
3	жилищные услуги: вода, электроэнергия, газ и другие виды топлива	0,227260	ДСХ
....	
75	почтовые услуги	1,000000	наивный
76	услуги телефонной и факсимильной связи	1,000000	наивный
77	прочие услуги, не отнесенные к другим категориям	1,000000	наивный

В результате анализа выяснилось, что для большинства категорий метод двойного сглаживания Холта показал наилучшие результаты – так случилось для 65 из 77 категорий. Для 7 категорий лучшим оказался наивный одношаговый прогноз, для 4 категорий – ARIMA(2,1,2), для 1 категории – метод двойного сглаживания Брауна и ни для одной категории – метод простого экспоненциального сглаживания.

Помимо проверки качества работы методов для последних 5 месяцев, была составлена также таблица сравнения методов прогнозирования для всего временного интервала с 2003 до 2020 года, результаты сравнения приведены в таблице 7.

В отличие от прогнозирования на 5 последних месяцев, в данном случае наилучшими результатами оказались у метода ARIMA(2,1,2). Среди всех 77 категорий метод ARIMA(2,1,2) оказался лучшим для 29 категорий, метод двойного сглаживания Холта – для 24 категорий, наивный одношаговый прогноз – для 21 категории, метод двойного сглаживания Брауна – для 3 категорий, метод простого экспоненциального сглаживания – ни для одной категории.

Таблица 7.

Категории товаров и услуг, проранжированные по значениям ошибки MASE для различных методов прогнозирования за период 2003-2020

№	Категория	мин. ошибка	Метод
1	молочные изделия, сыр и яйца	0,511349	ARIMA
2	одежда и обувь	0,566484	ARIMA
3	одежда	0,612997	ARIMA
....	
75	почтовые услуги	1,000000	наивный

76	услуги телефонной и факсимильной связи	1,000000	наивный
77	прочие услуги, не отнесенные к другим категориям	1,000000	наивный

Заключение

После проведения анализа результатов прогнозирования ценовых рядов по 77 категориям за период 2003–2020 годов, для каждой из категорий были выявлены методы, которые дают прогнозы с наибольшей точностью. Для некоторых категорий товаров и услуг один и тот же метод оказывался наиболее эффективным и на протяжении всего временного отрезка (период с 2003 по 2020 годы), и при прогнозировании для последних 5 месяцев (период с сентября 2019 по январь 2020 года). Для подобных категорий товаров и услуг можно сделать вывод о том, что один и тот же метод постоянно оказывается наиболее эффективным для различных временных интервалов. В то же время, для некоторых категорий товаров и услуг, наиболее эффективный метод мог меняться в зависимости от временного интервала – это означает, что однозначно определить наилучший метод, в отличие от предыдущего случая, для данных категорий нельзя.

Наиболее точные прогнозы в среднем были получены при использовании методов ARIMA и двойного сглаживания Холта. Методы простого экспоненциального сглаживания и двойного сглаживания Брауна оказались значительно менее эффективными. Метод наивного одношагового прогнозирования оказался эффективным лишь для определённой категории товаров – чаще всего это оказывались товары и услуги, цены на которые образуются не на рынке с совершенной конкуренцией («газ», «связь», «водоснабжение»).

Полученные результаты позволяют осуществлять сравнительно точное краткосрочное прогнозирование по всем основным категориям товаров и услуг в Кыргызской Республике. Анализ эффективности различных методов прогнозирования даёт возможность сразу использовать наиболее релевантные методы, а также определить, существует ли возможность качественно прогнозировать выбранную категорию товаров, или услуг.

Список литературы

1. Asteriou, D. ARIMA Models and the Box–Jenkins Methodology / D. Asteriou, S.G. Hall. - Applied Econometrics, volume 2, 2011. - 265-286 p.
2. Brown, R.G. Exponential Smoothing for Predicting Demand / R.G. Brown. - Cambridge, Massachusetts, 1956. – 15 p.
3. Holt, C.C. Forecasting Trends and Seasonal by Exponentially Weighted Averages / C.C. Holt. - Office of Naval Research Memorandum, volume 52, 1957. – 5 p.
4. Hyndman, R.J. Another look at measures of forecast accuracy / R.J. Hyndman, A.B. Koehler. – International Journal of Forecasting, volume 22 issue 5, 2006. – 679-688 p.
5. Makridakis, S. Statistical and Machine Learning forecasting methods: Concerns and ways forward / S. Makridakis, E. Spiliotis, V. Assimakopoulos. – Plos One, volume 13 issue 3, 2018. – 26 p.
6. Makridakis, S. The M4 Competition: Results, findings, conclusion and way forward / S. Makridakis, E. Spiliotis, V. Assimakopoulos. - International Journal of Forecasting, volume 34 issue 4, 2018. – 802-808 p.
7. Национальный статистический комитет. Цены и тарифы.
8. URL: <http://www.stat.kg/ru/statistics/ceny-i-tarify/>