

## ИСПОЛЬЗОВАНИЕ АНАЛИТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗА ПРОДАЖИ КВАРТИР НА ПРИМЕРЕ ЖК «АКАДЕМИЯ»

© Сафронова С.Э., Чеховская С.А., 2015

Иркутский государственный университет, г. Иркутск

В настоящее время, в условиях высокой неопределенности бизнес-прогнозирование является полезным инструментом для принятия управленческих решений. Данная статья посвящена изучению различных аналитических моделей и оценки их эффективности, а также практических расчётов значений данных по продажам квадратных метров квартир и их прогнозирование на 2015 год. Ни одна бизнес-модель не позволяет сделать точный прогноз на длительный период времени, однако может быть полезна для представления общей картины и видения текущих перспектив.

*Ключевые слова:* бизнес-прогнозирование, RMSE, U-статистика.

Люди на протяжении всего своего существования хотели знать, что будет завтра. Зачастую это обусловлено простым интересом, но мы знаем немало других примеров, когда знание того, что ждет нас завтра, действительно важно. Самый простой пример, приходящий в голову, это рынок. Любой продавец хочет продать как можно больше своей продукции и получить максимальную прибыль. Для того чтобы добиться желаемого результата, ему необходимо знать, сколько продукции потребуется

завтра (чтобы подготовить нужное количество товара), сколько будет стоить его продукция (рынок изменчив, цены постоянно меняются) и так далее. Конечно, опытный продавец в силах прикинуть сам, чего ждать от следующего дня. Но в последние годы колебания цен и спроса потребителей заметно выросли. Экономическая ситуация нестабильна. В условиях высокой неопределенности очень трудно прийти к верному решению. Интуиция, на которую продавец мог раньше положиться, может подсказать неверное

решение. Более того, менеджерам приходится принимать ответственные решения, касающиеся немалых денежных средств. Если решение окажется ошибочным, то компания не просто потеряет значительную сумму денег, но также может лишиться клиентов, инвесторов и союзников. С целью облегчения принятия решений были созданы системные аналитические модели. Данные модели способны высчитать прогнозное значение. Существует огромное количество таких моделей. Одни лучше справляются с обычными данными, другие с данными, в которых присутствует тренд. Несмотря на разнообразие методов бизнес-прогнозирования, не стоит забывать, что ни одна из моделей не способна дать 100% точный прогноз.

Перед тем, как начать применять какую-либо модель, необходимо проанализировать данные или, другими словами, имеющийся временной ряд (ряд наблюдаемых значений, которые последовательно расположены в историческом порядке по возрастанию) на наличие тренда, сезонности, цикличности, шума или их отсутствия. Рассмотрим более подробно каждую из компонентов временного ряда. Трендовая компонента — линейная или нелинейная компонента, которая меняется с течением времени. Отражает возрастания и убывания данных на протяжении длительного периода времени. Сезонная компонента показывает временные колебания внутри года. Принято считать, что сезонность показывает изменения, повторяющиеся каждый год. Циклическая компонента определяет периодические колебания, границы которых превышают временные рамки в один год. Длина цикла определяется расстоянием между двумя ближайшими вершинами или впадинами в масштабах года. На графике цикличность изображается как волнообразные отклонения от линии тренда. Все вышеперечисленные компоненты являются систематическими и характеризуются тем, что могут в одно и то же время присутствовать во временном ряду. Последняя составляющая временного ряда — случайная ошибка (шум). Не влияет на ряд постоянно. Следует отметить, что она является несистематической компонентой. Основными причинами, из-за которых возникает шум, являются: факторы внезапного действия и действия текущих факторов. Шум, возникающий из-за второй причины, может быть результатом ошибочных наблюдений [1]. Главная цель анализа временного ряда — обнаружение и оформление количественного выражения каждой компоненты с целью использования полученной информации для получения прогнозных значений ряда. Если имеется несколько переменных, следует определить степень их зависимости между собой. При наличии нескольких независимых переменных требуется установить не только их зависимость с игроком, но и оценить степень зависимости между

самими иксами. В случае, когда наблюдается высокая взаимосвязь между независимыми переменными, возникает явление мультиколлинеарности. Необходимо будет провести дополнительное исследование данных и, возможно, избавиться от нежелательной переменной. После обработки данных мы приступаем к выбору модели. Для того, чтобы подобрать наиболее подходящую модель, нам следует просчитать несколько моделей и найти для каждой RMSE (сумма квадратов ошибок, поделенная на количество этих ошибок). Далее мы находим U-статистику. U-статистика показывает отношение RMSE выбранной модели к RMSE первой наивной модели, о которой мы расскажем позднее. Если данный показатель больше 1, то целесообразнее использовать НМ 1 и наоборот. При выборе лучшей для нас модели мы обращаем внимание именно на U-статистику: чем ниже показатель, тем качественнее будет прогноз. Перейдем к описанию моделей, которые будут использованы при обработке данных. Первая: наивная модель 1 и наивная модель 2. уже из названия мы видим, что методы вычисления прогнозных значений будут простыми. Предполагается, что будущее значение лучше всего описывается последним периодом временного ряда. Поэтому в моделях прогноз является очень простой функцией:  $F_t = A_{t-1}$  и  $F_t = A_{t-1} + p \cdot (A_{t-1} - A_{t-2})$  соответственно. Не стоит ждать большой точности от таких примитивных моделей. Вторая модель — скользящее среднее. Используется для сглаживания колебаний и выявления тенденций. Эффект сглаживания достигается за счет получения средних значений за предыдущий период. При расчетах каждое значение находится заново, так как учитывается ограниченное количество предыдущих данных, наблюдается «скольжение» по ряду. Экспоненциальное сглаживание — один из простых способов, позволяющих выровнять ряд данных. Учитывается коэффициент сглаживания — альфа, колеблющийся в пределах от 0 до 1. Если требуется придать вес более свежим значениям, то коэффициент берут ближе к 0, если «старые» данные важнее, коэффициент, соответственно, берется ближе к 1. Модель Хольта наилучшим образом подходит для данных с трендом. К преимуществам данного метода можно отнести возможность прогноза на длительный период. Это более сложная модель, в которой учитывается сглаженный экспоненциальный ряд и тренд. Модель Винтерса больше подходит к данным с сезонностью. Помимо экспоненциального ряда и тренда берут во внимание сезонность. при вычислениях используют 3 показателя: альфа — сглаживание уровня данных, бетта — сглаживание индекса сезонности, гамма — сглаживание тренда. При оптимизации модели каждый из показателей можно менять для получения улучшенных результатов. Адаптивная модель отличается тем, что коэффициент для расчетов не подбирается

вручную, а высчитывается с помощью формулы, поэтому для каждого нового значения находится свой коэффициент. Аддитивная и мультипликативная модели. Значительное количество случаев говорит о том, что весь временной ряд можно изобразить суммой или произведением трех компонент, таких как тренд, цикличность и сезонность. Первая, аддитивная модель будет представлена как сумм трех компонент, а мультипликативная — произведением [2].

Для моделирования вышеописанных аналитических моделей были выбраны данные по продажам квартир ЖК «Академия» в период с 2011 по 2014 года. Застройщиком данного жилищного комплекса выступала крупнейшая компания в Восточной Сибири ОАО «Сибавиастрой» [3].

В исторических данных за каждый месяц соответствующего года представлена сумма квадратных метров жилищной площади, которая была реализована участникам долевого строительства секционного жилищного комплекса «Академия». Соответственно, при расчете прогнозируемых значений конечным результатом будет являться потенциально продаваемая сумма площадей квартир за последующие периоды. Отталкиваясь от тех данных, которые получают авторы статьи, можно спрогнозировать выручку компании от реализации всех площадей. Стоит отметить также то, что все квартиры, квадратные площади которых будут использоваться для расчётов, находятся в одном Эконом классе. Использование различных аналитических моделей позволят авторам выявить такие компоненты временного ряда, как тренд, цикличность, случайная ошибка (шум) и сезонности или доказать, что таковые отсутствуют. Данные прогноза позволят руководителю отдела продаж грамотно распределить нагрузку на менеджеров в зависимости от высокого или низкого сезона продаж; а также помогут высшему руководству в принятии управленческих и финансовых решений.

Теперь перейдем непосредственно к расчетам. Был проведен анализ точечного и интервального ряда данных. Результаты первого анализа: среднее значение — 921, мода — 0, медиана — 841, эксцесс — 0.28, асимметричность — 0.85. Итог интервального: среднее значение — 938, мода № 1 и № 2 — 402.18 (в данных присутствовало 2 моду, значения получились одинаковыми), медиана — 804. Опираясь на полученные результаты, можно сделать вывод о том, что наши данные не распределены по нормальному закону, так как среднее, мода и медиана не равны между собой; наблюдается ярко выраженное правостороннее распределение, так как асимметрия положительна и близка к единице; присутствует острокопечное распределение, потому что эксцесс положителен. На диаграммах, представленных ниже, вы можете видеть гистограмму частот ряда (Рисунок 1) и полигон относительных частот (Рисунок 2).



Рис. 1

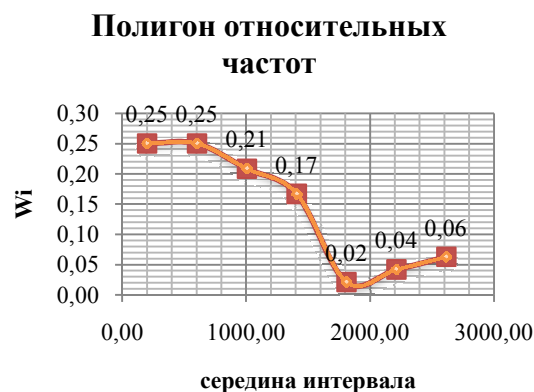


Рис 2

Коррелограмма и автокорреляционная функция говорят о том, что наш ряд данных случаен (коэффициенты автокорреляции близки к 0), следовательно, значения не имеют зависимости между собой. Также можно предположить, что данные имеют тренд: первые коэффициенты значительно отличаются от нуля, а далее следует затухание.

Наивная модель 1 показала нам  $RMSE = 515.72$ . Что ненамного отличается от наивной модели 2. Стоит заметить то, что наивная модель 2 позволяет учитывать наличие тренда, поэтому предполагается, что она будет давать более качественный результат. Однако её необходимо было оптимизировать, чтобы определить наилучшую степень влияние тренда. После оптимизации имеем  $RMSE = 446.11$ , а U-статистика равна 0.86. Данное значение меньше 1, следовательно, используем вторую наивную модель.

Следующая модель — скользящее среднее. Как и модель НМ2, эту модель так же необходимо оптимизировать. Мы получили такие результаты:  $RMSE$  модели — 393.6, U-статистика — 0.76.

Экспоненциальное сглаживание показало следующее (после оптимизации):  $RMSE = 510$ , U-статистика — 0.99.

Метод Хольта. Для данного метода были использованы коэффициенты альфа и бета, значения которых изначально были заданы нами самостоятельно, но позже были оптимизированы. В результате имеет  $RMSE = 695.38$ , а U — 1.35. В

качестве запускового значения использовали среднее значения ряда.

Модель Винтерса. Для запуска модели был взят дополнительные «запусковой» год, в качестве первого значения использовали не среднее значение ряда, а первое значение фактического ряда. В результате оптимизации модели были получены следующие данные: коэффициент альфа — 0.85, коэффициент бета — 0.000000000001, коэффициент гамма — 0.014, RMSE модели — 504, U-статистика — 0.98 (Рисунок 3).



Рис. 3

Адаптивная модель показала нам следующие результаты: RMSE — 555, 65 при бета 0.3. U-статистика — 1.077. По ниже представленному графику (Рисунок 4) мы можем заметить то, что она является не самой точной для дальнейшего прогноза.



Рис. 4

Приступим к декомпозиции временного ряда:

Аддитивная модель, как уже было сказано ранее, представляет временной ряд суммой трендовой циклической и сезонной компонент. Для проведения расчетов была выполнена регрессия: коэффициент для Y — 1 587, для X — минус 27.2. SSE или сумма квадратов ошибок — 15 925 494.6; SST или сумма квадратов разностей фактического и среднего значений — 23 968 280; RMSE — 576, U-статистика — 1.12. На графике, приведенном ниже (Рисунок 5), мы можем видеть фактические данные, выделенные синим цветом, красные прогнозные значения, и линию тренда, черная прямая.



Рис. 5

Мультипликативная модель представляет временной ряд как произведение трех систематических компонент. Коэффициенты, взятые из регрессии: для Y — 1 774, для X — минус 33. Сумма квадратов ошибок или SSE — 19 096 937, и SST — 23 968 280, RMSE модели — 630 и U-статистика 1.22. Ниже представлен график мультипликативной модели (Рисунок 6). Синей линией обозначены фактические данные, красной — прогнозные значения (Tt\*St), черной — линия тренда.



Рис. 6

Для наглядности выполним сведение полученных RMSE и U в таблицу:

Модель	RMSE	U
HM1	515.72	1.00
HM2	446.11	0.87
Среднее скользящее	393.6	0.76
Экспоненциальное сглаживание	510.00	0.99
Модель Хольта	695.38	1.35
Модель Винтерса	503.60	0.98
Адаптивная модель	555.65	1.08
Аддитивная модель	576	1.12
Мультипликативная модель	630	1.22

Из таблицы видно, что наименьшее RMSE у модели скользящего среднего. Следовательно, эта модель является наилучшей для получения прогнозных значений нашего временного ряда. На основе полученных данных, мы делаем прогноз. Прогнозное значение на следующий месяц равно 69.09 (Рисунок 7).



Рис. 7

В данной статье мы рассмотрели несколько видов моделей бизнес-прогнозирования, а именно: НМ1, НМ2, среднее скользящее, экспоненциальное сглаживание, модель Хольта, модель Винтерса, адаптивную, аддитивную и мультипликативную модели. Для каждой модели были вычислены RMSE и U-статистика. На основе полученных результатов была выбрана наиболее подходящая нам модель, после чего было найдено прогнозное значение. ■

---

1. URL: <http://reftrend.ru/792081.html>

2. Ханк Д. Э. Бизнес-прогнозирование: научно-популярное издание / Д. Э. Ханк, Д. У. Уичерн, А. Дж. Райте —М.: Издательский дом «Вильямс», 2003. — 651 с.

3. URL: <http://sibaviastroy.ru/>

### СПИСОК ЛИТЕРАТУРЫ

URL: <http://reftrend.ru/792081.html>

URL: <http://sibaviastroy.ru/>

Ханк Д. Э. Бизнес-прогнозирование: научно-популярное издание / Д. Э. Ханк, Д. У. Уичерн, А. Дж. Райте —М.: Издательский дом «Вильямс», 2003. — 651 с.

---

## The use of analytical models for the forecast of department's sales based on the data of housing complex «Akademija»

© Safronova S., Chekhovskaya S., 2015

At the present days, at the time of a high uncertainty, business forecasting is one of the useful tools for decision making in management. The article is devoted to the study of various analytical models and to access of their effectiveness. Also, the practical calculations will be included. None of the business models allow to make accurate prediction for a long period of time, but can be useful for an overall picture and vision of the current prospects.

*Keywords:* business forecasting, RMSE, U – statistic