

---

---

# Методы и устройства повышения качества передачи информации

---

---

УДК 519.711.2

## Сравнение методов прогнозирования для разработки интеллектуального модуля системы раннего предупреждения об ошибках на производстве

Киселева В.А.

В статье рассмотрены методы прогнозирования временных рядов: линейная регрессия, экспоненциальное сглаживание, скользящее среднее и взвешенное скользящее среднее. Исходные данные для построения моделей были собраны на производстве, конкретно в Колесопрокатном цехе АО «ВМЗ». В работе были построены четыре модели на основе рассмотренных методов и произведена их оценка при помощи коэффициента детерминации. Целью данной статьи является нахождение эффективного способа прогнозирования.

*Ключевые слова:* прогнозирование, временные ряды, регрессия, сглаживание, фильтр Хэмплера, коэффициент детерминации, коэффициент корреляции.

### Введение

Хранилище данных – класс систем в IT, в которых данные разной природы (логистика, производство и другое) с точки зрения бизнес-процессов собираются из многих источников для построения различных аналитик. Одна из задач хранилища – это обеспечение качества данных. Информация загружается из систем-источников, в которых могут быть ошибки, наличие которых неважно для производственного процесса, но при построении аналитики важна корректность и размеченность данных. Производственных процессов большое количество, каждый из которых знать невозможно, но нужно масштабно оценивать качество данных. Здесь стоит воспользоваться математическим прогнозированием, на основании которого можно будет делать выводы о качестве данных.

Математически прогнозировать ту или иную величину, например объем выручки или объем производства, следует на основании исторических периодов, которые уже были выверены, в качестве данных которых есть уверенность, вычислять прогноз на следующий период. Это позволит сравнить факт с теорией, и в случае отклонения заранее узнать о наличии потенциальных ошибок.

Следовательно, необходимо разработать систему, которая позволит узнать об ошибке в процессе загрузки данных до составления аналитики по ним. Сложность в том, что исторический ряд короткий, в основном величины характеризуются месяцем, а данные старше трёх лет нерелевантны, так как экономика нестабильна.

Исходные наблюдения характеризуются временным периодом, следовательно, для решения данной задачи стоит применять алгоритмы прогнозирования для временных рядов. Наиболее точный результат в данном случае обеспечат методы экстраполяции, к которым относится: экспоненциальное сглаживание, скользящее среднее и взвешенное скользящее среднее.

### Временные ряды

Временной ряд – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления. Он состоит из наблюдений, сделанных в разные моменты времени, и может быть представлен в виде графика или таблицы. Временные ряды используются для анализа и прогнозирования

изменений во времени, таких как экономические показатели, финансовые данные, климатические условия и другие явления. Анализ временных рядов включает в себя идентификацию трендов, сезонных колебаний, цикличности и случайных флуктуаций в данных.

У временных рядов есть несколько основных свойств:

1. Периодичность: временные ряды могут быть периодическими или аperiodическими. Периодические временные ряды имеют значения, которые повторяются с определенной периодичностью, например, ежедневные данные о температуре. Аperiodические временные ряды не имеют явной периодичности, например, ежедневные данные о финансовых индексах.

2. Тренд: временные ряды могут иметь тренд, который представляет собой долгосрочное изменение значения ряда во времени. Тренд может быть восходящим (значения растут), нисходящим (значения падают) или горизонтальным (значения остаются стабильными).

3. Сезонность: временные ряды могут иметь сезонность, которая представляет собой повторяющиеся паттерны или циклы в данных. Сезонность может быть годовой (например, увеличение продаж во время праздников), месячной (например, увеличение продаж в начале каждого месяца) или иной периодичности.

4. Шум: временные ряды могут содержать случайные или непредсказуемые флуктуации, которые называются шумом или случайной компонентой. Шум может быть вызван случайными событиями или ошибками измерений.

5. Автокорреляция: временные ряды могут иметь автокорреляцию, которая представляет собой связь между значениями ряда в разные моменты времени. Автокорреляция может быть положительной (значения взаимосвязаны и следуют за другими) или отрицательной (значения взаимосвязаны и противоположны друг другу).

6. Стационарность: временные ряды могут быть стационарными или нестационарными.

Стационарный временной ряд имеет постоянное среднее значение и постоянную дисперсию во времени, а также отсутствие систематического тренда или сезонности. Нестационарные временные ряды могут иметь изменяющееся среднее значение, дисперсию или тренд.

7. Цикличность: временные ряды могут иметь цикличность, которая представляет собой повторяющиеся волны или колебания в данных. Цикличность может быть связана с экономическими циклами или другими долгосрочными факторами.

Эти свойства временных рядов могут быть учтены для прогнозирования временных рядов.

Существует несколько методов прогнозирования временных рядов, включая:

1. Метод сглаживания: этот метод основан на предположении, что будущие значения временного ряда будут подобным образом сглаживаться относительно прошлых значений. Примеры методов сглаживания включают метод скользящего среднего и экспоненциальное сглаживание.

2. Методы регрессии: эти методы используются для прогнозирования временных рядов, основываясь на связи между временным рядом и другими переменными, которые могут влиять на его поведение. Например, можно использовать линейную регрессию или авторегрессионные модели для прогнозирования временных рядов.

3. Авторегрессионные интегрированные скользящие средние (ARIMA): это одна из наиболее распространенных методов прогнозирования временных рядов. ARIMA-модель используется для моделирования тренда, сезонности и случайных флуктуаций в данных.

4. Методы машинного обучения: с развитием машинного обучения стали доступны новые методы прогнозирования временных рядов, такие как нейронные сети и градиентный бустинг. Эти методы могут быть более гибкими и точными, но требуют большего количества данных и вычислительных ресурсов.

5. Сезонные модели: эти модели используются для прогнозирования сезонных колебаний во временных рядах. Они могут включать в себя методы сглаживания или регрессии, а также дополнительные компоненты, учитывающие сезонность.

Каждый метод имеет свои преимущества и ограничения, и выбор метода зависит от характеристик временного ряда и целей прогнозирования. Конкретно для решения поставленной задачи, подходят только методы сглаживания или регрессии, так как исторический ряд экспериментальных данных короткий, а именно ограничен 30 наблюдениями.

### Обзор научных статей

В статье [1] представлены результаты прогнозирования уровня иностранных инвестиций в российской экономике по модели экспоненциального сглаживания. Модель построена при помощи статистического программного продукта Statistica, в котором автоматически была найдена оптимальная константа сглаживания. Также автор сообщает о том, что при прогнозировании не учитываются качественные характеристики, которые могут повлиять на результаты прогнозирования, например, введенные санкции против РФ. Построенная модель оценивается шестью показателями: средняя ошибка, средняя абсолютная ошибка, сумма квадратов ошибок, относительная ошибка, средняя относительная ошибка, средняя абсолютная относительная ошибка.

В работе [2] рассмотрен метод оценки прогнозируемого значения случайного процесса, отсчеты которого коррелированы. В задачах, когда объем статистических данных ограничен, наиболее применимыми оказываются экстраполяционные методы прогнозирования, к которым относятся метод наименьших квадратов и совокупность методов сглаживания. Здесь проанализирована возможность применения метода наименьших квадратов для прогнозирования

в условиях малой выборки, например, случайного процесса спроса, наблюдения которого коррелированы.

В статье [3] рассматривается метод взвешенной скользящей средней для прогнозирования временных рядов, вычисляются необходимые параметры и строится прогноз по данным конкретного примера. Метод скользящего среднего применяется для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов. Автор ставит перед собой задачу – предсказать энергозатраты на будущее с минимальной погрешностью. При построении прогноза, последние значения экспериментальной выборки считаются более значимыми, они являются более актуальными и являются весомее предыдущих. В результате был построен прогноз значения на 1 день с погрешностью отклонения менее 1%. Автор делает вывод о том, что стоит учитывать особенности данных и для каждого конкретного случая необходимо подбирать период и коэффициенты.

В статье [4] рассматриваются модели и методы прогнозирования временных рядов, описываются их достоинства и недостатки. Автор описывает временной ряд как непрерывное явление, протекающее во времени, но информация об этом явлении поступает выборочно. Временные ряды могут характеризоваться по группам: время, интервал, форма представления уровней, содержание показателей; но зачастую для хорошей прогностической модели выбирают полный временной ряд с относительными величинами.

В работе [5] описывается метод нахождения и сглаживания выбросов во временном ряду при помощи фильтра Хэмплера, который в своей основе использует скользящее среднее с заданным окном. Для каждой итерации фильтр вычисляет медиану и стандартное отклонение, которое выражается в среднем абсолютном значении и обозначается как MAD. Метод имеет два настраиваемых параметра: размер раздвижного

окна и количество стандартных отклонений, которые идентифицируют выброс. Фильтр имеет ряд преимуществ: простота реализации, быстрота выполнения и применение как на малых, так и на больших объемах данных.

В статье [6] рассматривается проблема прогнозирования временных рядов с использованием различных методов. Автор проводит сравнительный анализ эффективности различных методов прогнозирования, таких как экспоненциальное сглаживание, авторегрессия, скользящее среднее и другие. Один из основных выводов статьи заключается в том, что выбор метода прогнозирования зависит от характера временного ряда и целей прогнозирования. Например, если временной ряд имеет линейную тенденцию, то лучше использовать метод экспоненциального сглаживания. Если же временной ряд имеет сезонность, то лучше использовать методы авторегрессии или скользящего среднего с учетом сезонности. В работе отмечено, что для повышения точности прогнозирования можно использовать комбинации различных методов, например, комбинацию экспоненциального сглаживания и авторегрессии.

Статья [7] посвящена исследованию эффективности использования различных модификаций метода скользящих средних для прогнозирования временных рядов. В статье отмечены недостатки скользящей средней при прогнозировании, основными из которых являются проблема «запаздывания» и чрезмерной чувствительности метода, что привело к использованию различных модификаций метода, так называемого «семейства скользящих средних». Одной из наиболее известных модификаций инструмента является взвешенная скользящая средняя, которая полагается, что последние значения исходной функции являются более значимыми, нежели предшествующие, причем функция значимости является линейно-убывающей.

### Линейная регрессия

Линейная регрессия строится на основе метода наименьших квадратов, который заключается в

нахождении оптимальных параметров моделей линейной регрессии, таких что сумма квадратов невязок регрессионных остатков минимальна. МНК предоставляет возможность, в ходе анализа экспериментальных данных, получить аналитическую функцию, которая проходит настолько близко к исходным точкам, насколько это осуществимо.

Так как проводимое исследование является однофакторным, линию регрессии можно описать уравнением:

$$y = a_0 + a_1x.$$

В общем случае, задачу можно сформулировать так: пусть в ходе некоторого эксперимента были получены статистические данные, необходимо построить аналитическую зависимость, которая наиболее точно опишет результаты эксперимента. Для построения параметров функции необходимо использовать метод наименьших квадратов. Суть МНК заключается в следующем, функцию необходимо подобрать таким образом, чтобы сумма квадратов отклонений измеренных значений была наименьшей от расчётных. Под отклонением понимается разность результата эксперимента и полученного значения функции для одного и того же фактора  $x$ . Для наглядности представим это в графическом виде.

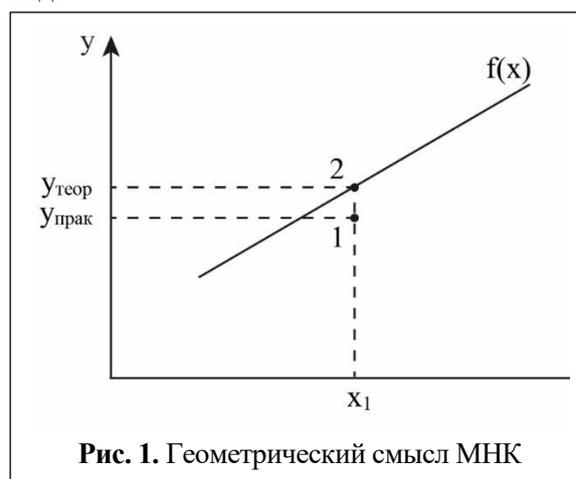


Рис. 1. Геометрический смысл МНК

Для оценки неизвестных параметров  $a_0$  и  $a_1$  линейной регрессии используется функция стоимости:

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2.$$

В статье [8] было проведено исследование применения метода регрессии при прогнозировании значений на этой же экспериментальной выборке, которое показало, что данные не имеют линейной зависимости, поэтому данный метод можно исключить при дальнейших исследованиях.

### Экспоненциальное сглаживание (ЕМА)

Метод экспоненциального сглаживания (метод Брауна) – это метод математического преобразования, используемый при прогнозировании временных рядов. Он применим при построении прогноза только на один период вперед.

Сглаженная кривая строится по формуле:

$$s_t = \begin{cases} y_t, & t = 1; \\ s_{t-1} + \alpha(y_t - s_{t-1}), & t > 1, \end{cases}$$

где  $s_t$  – сглаженный ряд,  $y_t$  – исходный ряд,  $\alpha$  – коэффициент сглаживания.

Прогнозное значение вычисляется по формуле:

$$s_{t+1} = \alpha \cdot y_t + (1 - \alpha)s_t.$$

Коэффициент сглаживания  $\alpha$  выбирается в пределах от 0 до 1, от его величины зависит, как быстро снижается вес влияния предшествующих наблюдений. Если  $\alpha$  близко к единице, то это приводит к учету при прогнозировании влияния последних наблюдений, и наоборот, если  $\alpha$  близко к нулю, то при прогнозировании учитываются все прошлые наблюдения.

### Скользящее среднее (МА)

Метод скользящего среднего представляет собой среднее арифметическое значение наблюдений в определенном окне или периоде, который "скользит" вдоль временного ряда.

Сглаживание с помощью скользящего среднего основано на том, что в средних величинах взаимно погашаются случайные отклонения, что происходит вследствие замены первоначальных уровней временного ряда средней

арифметической величиной внутри выбранного интервала времени. Полученное значение относится к середине выбранного интервала времени (периода).

Затем период сдвигается на одно наблюдение, и расчет средней повторяется. При этом периоды определения средней берутся все время одинаковыми. Таким образом, в каждом рассматриваемом случае средняя центрирована, т.е. отнесена к серединной точке интервала сглаживания и представляет собой уровень для этой точки.

При сглаживании временного ряда скользящими средними в расчетах участвуют все уровни ряда. Чем шире интервал сглаживания, тем более плавным получается тренд. Сглаженный ряд короче первоначального на  $(n-1)$  наблюдений, где  $n$  – величина интервала сглаживания.

При больших значениях  $n$  колеблемость сглаженного ряда значительно снижается. Одновременно заметно сокращается количество наблюдений, что создает трудности.

Выбор интервала сглаживания зависит от целей исследования. При этом следует руководствоваться тем, в какой период времени происходит действие, а следовательно, и устранение влияния случайных факторов.

Формула вычисления, скользящего среднего [1]

$$s_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}.$$

Прогнозное значение вычисляется по формуле

$$s_{t+1} = s_{t-1} + \frac{1}{n}(y_t - y_{t-1}), n = 3,$$

где  $s_{t-1}$  – скользящая средняя за два периода до прогнозного,  $n$  – это число уровней входящих в интервал сглаживания.

### Взвешенное скользящее среднее (WMA)

Метод взвешенного скользящего среднего – это алгоритм, вычисляющий прогноз для новых

наблюдений, на основе временных рядов, которые не имеют тренда. Веса присваиваются по возрастанию, то есть последнее наблюдение имеет больший вес, и наоборот.

Формула вычисления взвешенного скользящего среднего

$$s_t = \frac{2}{n(n+1)} \sum_{i=0}^{n-1} (n-i)y_{t-i},$$

где  $y_{t-i}$  – значение исходного наблюдения в момент времени, отдалённый от текущего на  $i$  интервалов,  $n$  – количество исходных наблюдений, для расчета скользящего среднего.

Прогнозное значение вычисляется по формуле

$$s_{t+1} = \frac{\sum_{i=t-n}^t (y_i \cdot w_i)}{\sum_{i=1}^n w_i},$$

где  $w_i$  – это вес  $i$  наблюдения,  $n$  – количество наблюдений,  $y_i$  – исходное наблюдение.

Предварительно, необходимо найти оптимальное количество наблюдений, на основе которых будет строиться прогноз. Сделать это можно при помощи среднеквадратичной ошибки

$$MSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}.$$

### Фильтр Хэмплера

Фильтр Хэмплера применяется для нахождения и замены выбросов в заданном временном ряду, который в своей основе использует скользящее среднее с заданным окном. Для каждой итерации вычисляется медиана и стандартное отклонение

$$MAD = \text{median}(|y_i - \bar{y}|).$$

Для того, чтобы MAD стал последовательной оценкой стандартного отклонения необходимо умножить его на постоянный коэффициент  $k=1,4826$ , предполагая, что данные подчиняются распределению Гаусса.

Если значение MAD больше, чем 3 стандартных отклонения, то это выброс, который меняется на скользящую медиану.

### Оценка полученных результатов

Оценка моделей осуществляется при помощи коэффициента детерминации и средней абсолютной ошибки в процентах.

Коэффициент детерминации

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

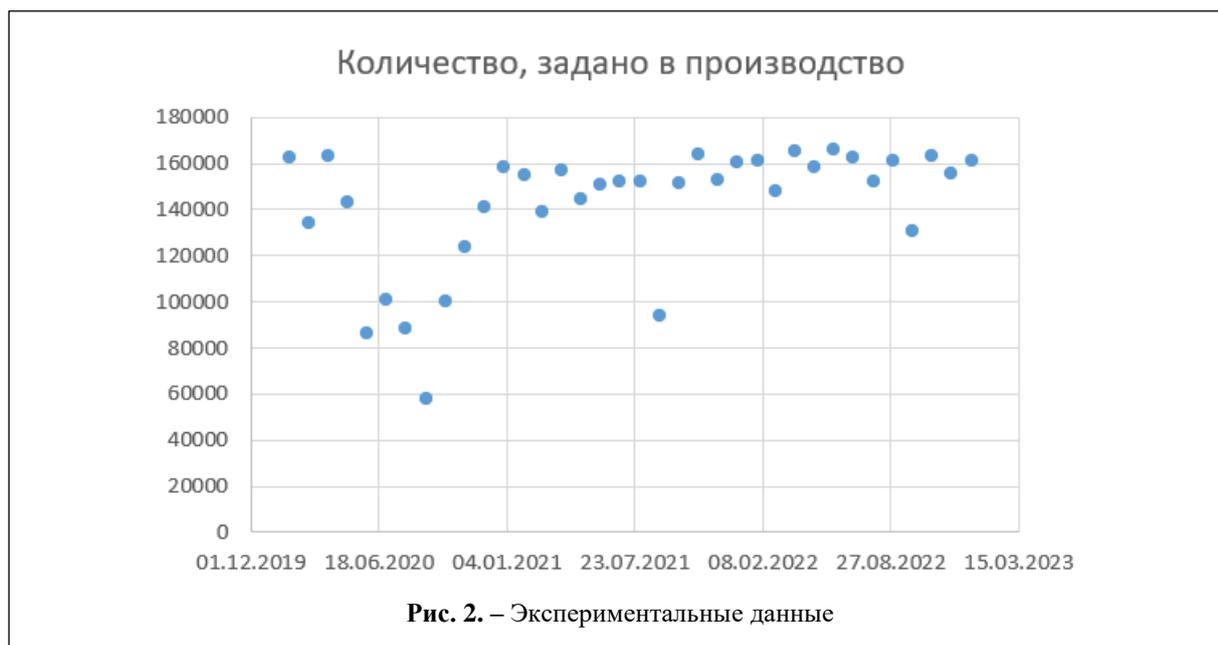


Рис. 2. – Экспериментальные данные

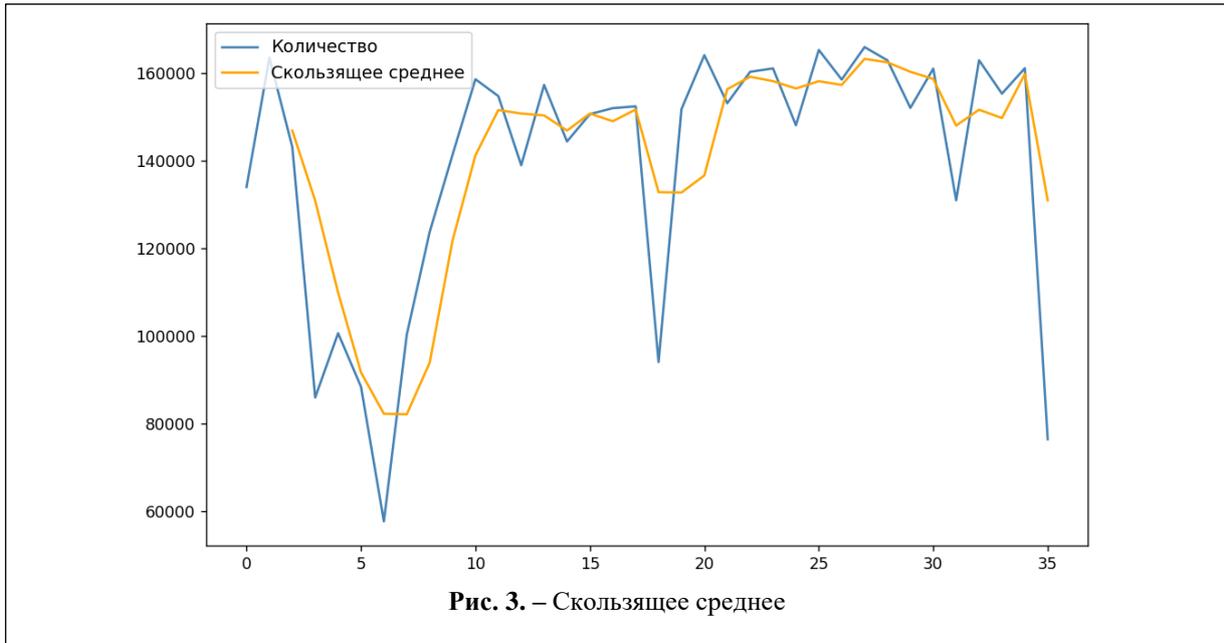


Рис. 3. – Скользящее среднее

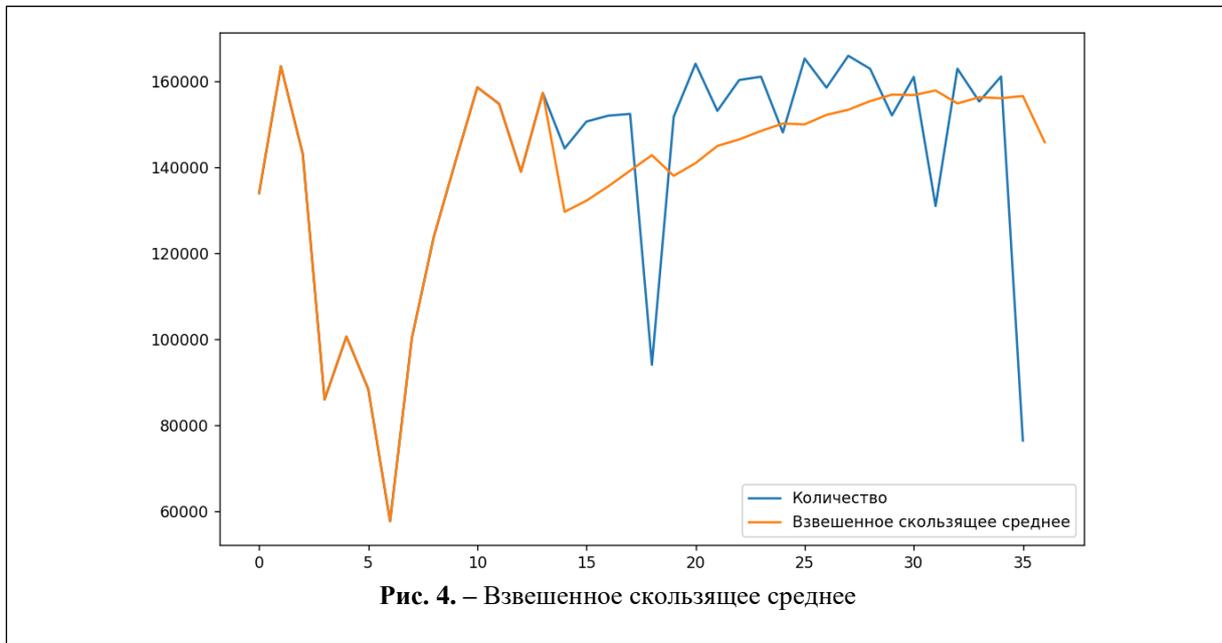


Рис. 4. – Взвешенное скользящее среднее

показывает наличие связи между факторами регрессии и зависимой переменной. Чем ближе коэффициент детерминации к единице, тем информативнее построенная модель.

Средняя абсолютная ошибка в процентах

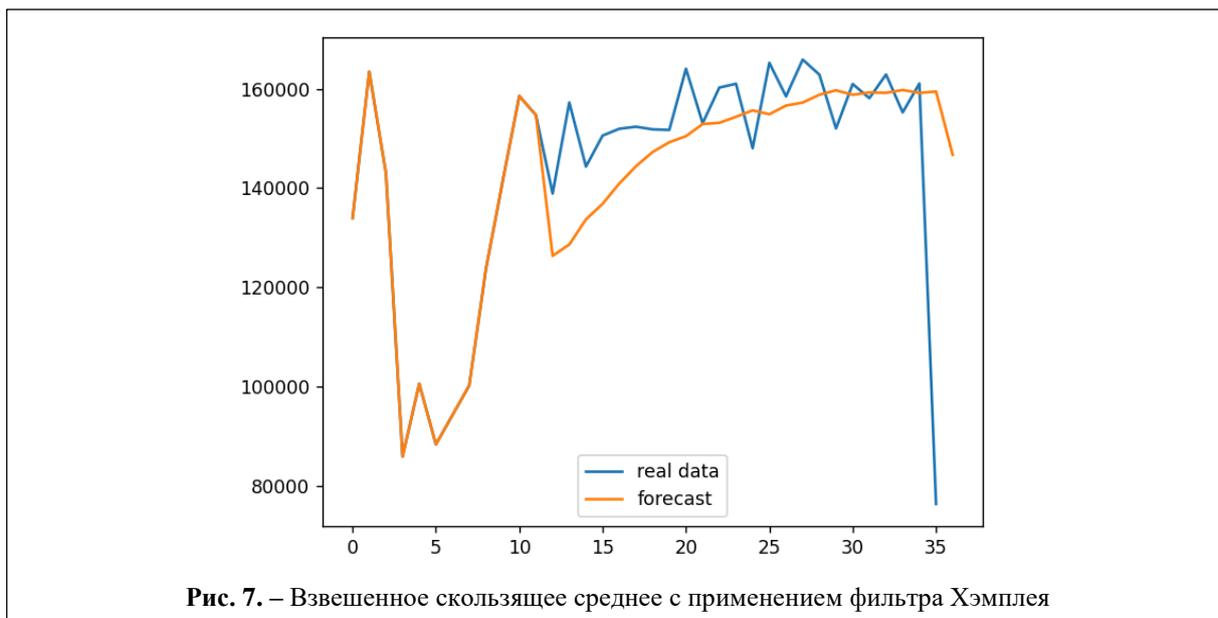
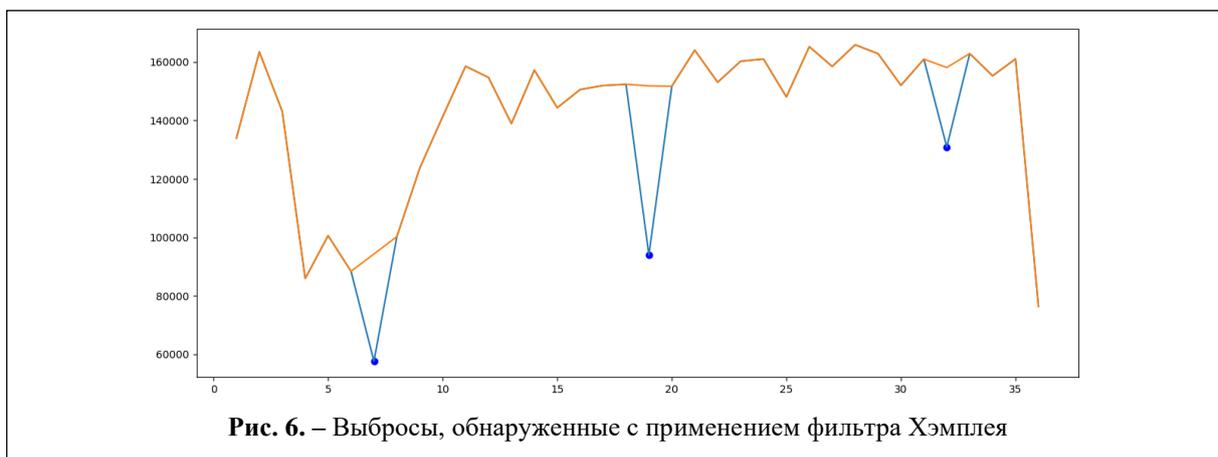
$$MAPE = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \cdot 100\%,$$

показывает среднюю разницу в процентах между фактическими и прогнозируемыми значениями.

### Реализация и сравнение методов прогнозирования

В ходе исследования были собраны данные по показателю «Количество, задано в производстве» функциональной отчетности по цеху КПЦ АО «ВМЗ».

Результаты эксперимента неоднозначные, наиболее сильную связь между фактом и прогнозом показывает метод взвешенного скользящего среднего с применением фильтра Хэмплия, а наименьшую среднюю абсолютную



ошибку в процентах показывает экспоненциальное сглаживание. Очевидно, что коэффици-

ент детерминации имеет больший вес для моделей прогнозирования. Следовательно, можно сделать вывод, что наилучший результат прогнозирования производственных показателей представляет метод взвешенного скользящего среднего с применением фильтра Хэмплера.

Результаты экспериментов приведены в табл. 1.

Метод	Истинное значение на 28.02.2023	Прогнозное значение на 28.02.2023	MAPE	R <sup>2</sup>
Экспоненциальное сглаживание	146587,971	117116.446	8.46%	0.7927
Скользящее среднее	146587,971	131511.955	23.08%	0.6175
Взвешенное скользящее среднее	146587,971	145786.682	8.49%	0.7701
Взвешенное скользящее среднее с применением фильтра Хэмплера	146587,971	146701.271	9.96%	0.8835

### Заключение

В результате исследования были построены четыре модели на основе рассмотренных методов и произведена их оценка при помощи коэффициента детерминации и средней абсолютной ошибки в процентах. Цель, поставленная при написании статьи, была выполнена, метод взвешенного скользящего среднего с применением фильтра Хэмплера показал наиболее эффективные результаты прогнозирования.

### Литература

1. Калюжин, Д. Г. Прогнозирование уровня иностранных инвестиций в российскую экономику по модели экспоненциального сглаживания / Д. Г. Калюжин // Аллея науки. – 2018. – Т. 4, № 10(26). – С. 568-571. – EDN YSNWIX.
2. Головкин, В. А. Прогнозирование коррелированного временного ряда по малой выборке исходных данных / В. А. Головкин, Я. Хазим. — Текст : непосредственный // Вестник НТУ "ХПИ". — 2014. — № 35(1078). — С. 43-47.

3. Снурницын, К. В. Прогнозирование временных рядов энергозатрат методом взвешенной скользящей средней / К. В. Снурницын. — Текст : непосредственный // Молодежь и XXI век - 2016. — Курск : Университетская книга, 2016. — С. 117-120.

4. Иванюк, В. А. Разработка метода прогнозирования на основе объединения трендов временного ряда / В. А. Иванюк // Управление разви-

тием крупномасштабных систем MLSD'2019 : Материалы двенадцатой международной конференции Научное электронное издание, Москва, 01–03 октября 2019 года / Под общей ред. С.Н. Васильева, А.Д. Цвиркуна. – Москва: Институт проблем управления им. В.А. Трапезникова РАН, 2019. – С. 519-521. – DOI 10.25728/mlsd.2019.1.0519. – EDN RFBYEL.

5. Generalized Hampel Filters / R. K. Pearson, Y. Neuvo, Ja. Astola, M. Gabbouj // Eurasip Journal on Advances in Signal Processing. – 2016. – Vol. 2016, No. 1. – P. 1-18. – DOI 10.1186/s13634-016-0383-6. – EDN BOGRUL.

6. Колодко, Д. В. Мониторинг валютного рынка Forex с помощью различных типов скользящих средних / Д. В. Колодко // Управление экономическими системами: электронный научный журнал. – 2013. – № 1(49). – С. 17. – EDN PXGGWH.

7. Шилов, А. В. Анализ достоинств и недостатков модифицированных скользящих средних при выработке прогнозных решений / А. В. Шилов // Приоритетные научные направления: от теории к практике. – 2015. – № 19. – С. 83-88. – EDN UNFHZT.

8. Киселева, В. А. Метод линейной регрессии для интеллектуального модуля системы раннего предупреждения ошибок на производстве / В. А. Киселева [Текст] // Международная научно-техническая конференция молодых ученых БГТУ им. В.Г. Шухова, посвященная 170-летию со дня рождения В.Г. Шухова. — Белгород: Белгородский государственный технологический университет (БГТУ) им. В.Г. Шухова, 2023. — С. 198-202.

9. Киселева, В. А., Рыжкова, М. Н. Обзор методов анализа малой выборки / В. А. Киселева, М. Н. Рыжкова [Текст] // XV Всероссийские научные Зворыкинские чтения 2023. — Муром: Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых», 2023. — С. 483-484.

10. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. М.: Финансы и статистика, 1987. - 230 с.

11. Булашев, С. В. Статистика для трейдеров / С. В. Булашев. — Москва : Компания Спутник+, 2003. — 245 с. — Текст : непосредственный.

12. Вайс, А. А. Краткосрочное прогнозирование таксационных показателей методом экспоненциального сглаживания / А. А. Вайс, А. С. Грицких. — Текст : непосредственный // Вестник современных исследований. — 2018. — № 5.1(20). — С. 315-317.

13. Арзамасцева, Э. А. Прогнозирование объёма реализации кирпичей методом экспоненциального сглаживания / Э. А. Арзамасцева. — Текст: непосредственный // Современная наука:

**Поступила 22 сентября 2023 г.**

The article discusses the methods of forecasting time series: linear regression, exponential smoothing, moving average and weighted moving average. The initial data for the construction of models were collected in production, specifically in the Wheel-rolling shop of JSC "VMZ". In this paper, four models were constructed based on the methods considered and their evaluation was performed using the coefficient of determination. The purpose of this article is to find an effective way of forecasting.

*Key words:* forecasting, time series, regression, smoothing, Hampel filter, determination coefficient, correlation coefficient.

*Киселева Владислава Андреевна* – магистрант 2 курса направления "Прикладная математика и информатика" Муромского института (филиала) ФГБОУ ВО «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

*Адрес:* 602264, г. Муром, ул. Орловская, д. 23.

теоретический и практический взгляд. — Уфа : Аэтерна, 2015. — С. 55-57.

14. Попов, Н.С. О корректности статистических выводов на основе использования коэффициента корреляции / Н. С. Попов, А. А. Баламутова, О. С. Филимонова, Л. Н. Чуксина. — Текст : непосредственный // Статистические методы исследования социально-экономических и экологических систем региона. — Тамбов : Издательский центр ФГБОУ ВО «ТГТУ», 2020. — С. 305-308.

15. Мхитарян В.С., Трошин Л.И., Корнилов И.А., Адамова Е.В., Шевченко К.К., Бамбаева Н.Я. Теория вероятностей и математическая статистика / Московский международный институт эконометрики, информатики, финансов и права. - М.: 2004. - 147с

16. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. – 6 изд., перераб. и доп. – М.: Дело, 2004. – 576 с.

17. Саммерфилд М. Программирование на Python 3 Подробное руководство. – Пер. с англ. – СПб.: Символ Плюс, 2009 – 608 с., ил. ISBN: 978 5 93286 161 5

18. Абдрахманов М.И. Python. Визуализация данных. Matplotlib. Seaborn. Mayavi., 2020. - 412 с.

19. Плас Джейк Вандер. Python для сложных задач. Наука о данных и машинное обучение. Руководство. / Плас Джейк Вандер. – М.: Питер, 2018.

20. Иванюк, В. А. Решение эконометрических задач на Python / В. А. Иванюк // Современная математика и концепции инновационного математического образования. – 2023. – Т. 10, № 1. – С. 229-233. – DOI 10.54965/24129895\_2023\_10\_1\_229. – EDN FVFNXB.