

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«АЛТАЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет математики и информационных технологий  
Кафедра теоретической кибернетики и прикладной математики

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ОЦЕНКИ КАЧЕСТВА  
ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ  
(магистерская диссертация)

Выполнила:  
студентка 458-М группы,  
Чернобаева Анна Александровна

\_\_\_\_\_  
(подпись)

Научный руководитель:  
д.ф.-м. н., профессор  
Алгазин Геннадий Иванович

\_\_\_\_\_  
(подпись)

Работа защищена:  
«\_\_»\_\_\_\_\_ 2017 г.

Оценка: \_\_\_\_\_

Председатель ГЭК:

\_\_\_\_\_  
(подпись)

\_\_\_\_\_  
Ф.И.О.

Допустить к защите:  
Зав. кафедрой, к.т.н., профессор  
Хворова Любовь Анатольевна

\_\_\_\_\_  
(подпись)

«\_\_»\_\_\_\_\_ 2017 г.

Барнаул 2017

## РЕФЕРАТ

Тема магистерской диссертации: «Информационные технологии оценки качества прогнозирования временных рядов»

Цель работы – провести сравнительный анализ оценок ошибок прогнозирования временных рядов с применением современных ИКТ и разработать рекомендации к выбору оценки качества прогнозирования на примере данных о численности нетрудоустроенных граждан Алтайского края.

Объект исследования – временные ряды.

Предмет исследования – оценки ошибок прогнозирования.

В результате магистерского исследования решены следующие задачи: осуществлен обзор моделей и методов прогнозирования временных рядов, выявлены достоинства и недостатки основных методов; проведен сравнительный анализ информационных технологий, используемых для прогнозирования временных рядов; построен временной ряд численности нетрудоустроенных граждан Алтайского края, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года; проведен расчет оценок ошибок прогнозирования для построенного временного ряда; проверены предложенные в литературе рекомендации по выбору по выбору оценок прогнозирования.

Объем работы 51 страниц, количество рисунков – 8, таблиц – 6, 27 использованных источников литературы.

Ключевые слова: моделирование, информационные технологии, прогнозирование, временные ряды, ошибка прогнозирования.

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	4
1. ПРОБЛЕМЫ ОЦЕНКИ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ ВРЕМЕННЫХ РЯДОВ..	7
1.1. Адекватность моделей прогнозирования и подходы к её оцениванию	7
1.2. Критерии оценки качества прогнозирования временных рядов.....	8
1.3. Ошибки прогнозирования .....	12
2. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРОГНОЗИРОВАНИИ И ОЦЕНИВАНИИ ОШИБОК МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ .....	13
2.1. Базовые модели временных рядов .....	13
2.2. Сравнительный анализ моделей прогнозирования. Достоинства и недостатки моделей .....	21
2.3. Информационные технологии и инструментарий оценивания прогнозирования .....	25
2.4. Систематизация ошибок моделей прогнозирования и методов их расчета.....	28
3. ОЦЕНКИ ОШИБОК ПРОГНОЗИРОВАНИЯ ПО МОДЕЛИ ПРОГНОЗИРОВАНИЯ ЧИСЛЕННОСТЬ НЕТРУДОУСТРОЕННЫХ ГРАЖДАН .....	34
3.1. Разработка модели прогнозирования с применением информационных технологий пакета MS EXCEL .....	34
3.2. Динамика и прогнозирование временного ряда численность нетрудоустроенных граждан Алтайского края.....	37
3.3. Расчеты оценок ошибок прогнозирования.....	40
3.4. Анализ приведенных в литературе рекомендаций к выбору оценок прогнозирования качества временных рядов.....	45
ЗАКЛЮЧЕНИЕ .....	47
БИБЛИОГРАФИЧЕСКИЙ СПИСОК .....	48

## ВВЕДЕНИЕ

На сегодняшний день полным ходом идет переход к рыночным отношениям в экономике и научно-технический прогресс чрезвычайно ускорили темпы внедрения во все сферы социально-экономической жизни общества последних научных разработок в области информационных технологий. Достижение Россией высоких результатов в экономике и завоевание места полноправного партнера в мировой экономической системе в значительной степени зависит от того, каковы будут масштабы использования современных информационных технологий во всех аспектах человеческой деятельности, а также от того, какую роль будут играть эти технологии в повышении эффективности экономических взаимоотношений.

Развитие прогностики как науки в последние десятилетия привело к созданию множества методов, процедур, приемов прогнозирования, неравноценных по своему значению. Согласно анализу иностранных и российских систематиков прогностики уже насчитывается более 100 способов прогнозирования, в связи, с чем перед специалистами возникает задача подбора способов, которые давали бы адекватные прогнозы для изучаемых процессов или систем. Строгие статистические теории о свойствах временных рядов ограничивают возможности классических способов прогнозирования.

С целью определения качества прогнозирования используются различные критерии: оценки ошибок прогнозирования, скорость расчета, интерпретируемость и другие. Анализ ошибки прогнозирования или точности прогноза считаются более значимыми при решении практических задач.

Как правило, в исследованиях применяют стандартные оценки ошибки прогнозирования, на основании которых формируется либо суждение о качестве предлагаемого механизма прогнозирования, либо осуществляется подбор оптимального механизма прогнозирования для нескольких объектов.

При решении подобных проблем в первую очередь используются «традиционные» критерии, характерные рассматриваемой данной предметной областью, несмотря на существующие на их минусы.

В данной работе приведены результаты анализа существующих и достаточно распространенных оценок ошибок прогнозирования, используемых в исследованиях по прогнозированию. Для каждой оценки приводятся формулы.

При исследовании временных рядов получить абсолютно точный прогноз почти невозможно, по этой причине значимой задачей считается оценка разных моделей прогнозирования с точки зрения определенных критериев качества. При этом надежность избранной модели прогнозирования оценивается периодическим сопоставлением фактических и прогнозируемых значений ряда.

Подобранная модель обязана соответствовать требованиям точности прогнозов, так как только в данном случае она может быть востребована. Критерии оценки точности прогнозирующей модели обязаны учитывать как разброс прогнозируемых сведений сравнительно с исходными значениями, так и разброс относительно тренда временного ряда.

Возможности информационных технологий применяются во многих отраслях науки, начиная от медицины и астрономии, завершая информатикой и экономикой. Между тем далеко не все потенциальные возможности методов прогнозирования изучены, но одними из их свойств являются возможности распознавания и классификации образов, работы с большими массивами данных, оценка прогноза, аппроксимация и выявление неочевидных зависимостей в данных временных рядов. На основе этого можно сделать вывод о существенном преимуществе их применения для анализа и прогнозирования динамик временных рядов.

Цель работы – провести сравнительный анализ оценок ошибок прогнозирования временных рядов с применением современных ИКТ и разработать рекомендации к выбору оценки качества прогнозирования на

примере данных о численности нетрудоустроенных граждан Алтайского края.

Для достижения цели были поставлены и решены следующие задачи:

1. Осуществить обзор моделей и методов прогнозирования временных рядов, выявить достоинства и недостатки основных методов;
2. провести сравнительный анализ информационных технологий, используемых для прогнозирования временных рядов;
3. построение временного ряда численности нетрудоустроенных граждан Алтайского края, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года ;
4. провести расчет оценок ошибок прогнозирования для построения временного ряда нетрудоустроенных граждан;
5. проверить предложенные в литературе рекомендации по выбору оценок прогнозирования.

Объект исследования – временные ряды.

Предмет исследования – оценки ошибок прогнозирования.

Методы исследования: при решении поставленных задач в работе использованы методы математического моделирования, построения и анализа временных рядов, современные ИКТ.

В работе получены следующие основные результаты: проведена апробация предложенных в литературе методов оценивания ошибок прогнозирования, на примере временного ряда нетрудоустроенных граждан Алтайского края.

# 1. ПРОБЛЕМЫ ОЦЕНКИ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ ВРЕМЕННЫХ РЯДОВ

## 1.1. Адекватность моделей прогнозирования и подходы к её оцениванию

Адекватность модели – это степень ее соответствия реальному объекту. Критерием адекватности модели служит только практическая деятельность, только эксперимент на реальном объекте.

Безмерно доверять точным математическим моделям далеко не всегда является эффективным. Каждая, сколь угодно сложная модель есть упрощение, отражающее настоящий уровень знаний о свойствах объекта прогнозирования.

Качество модели определяется ее адекватность исследуемому процессу и точностью.

Модель является адекватной, если систематическая компонента (тренд) определена верно. При правильном выборе вида тренда отклонения от него (т. е. разницы между фактическими уровнями ряда и их расчетными (вычисленными по модели) значениями) будут [11]:

1. носить произвольный характер. Это утверждение означает, что изменение остаточной компоненты не связано с изменением временного периода;
2. иметь равное нулю математическое ожидание;
3. нормально распределены;
4. независимы друг от друга (т. е. в остаточной последовательности не будет существенной автокорреляции).

Достоверность модели обуславливается степенью близости расчетных уровней ряда к фактическим данным, а также с помощью других характеристик.

Отклонения от тренда будут случайными, если выполнены следующие неравенства при 5%-ном уровне значимости [26]:

$$K_{max} < [3,3(\ln n + 1)] \quad (1.1)$$

$$V > \left[ \frac{1}{2}(n + 1 - 1,96\sqrt{n - 1}) \right] \quad (1.2)$$

Если хотя бы одно из неравенств не выполняется, то гипотеза о случайном характере отклонений уровней временного ряда от тренда с вероятностью 0,95 отвергается [9].

Проверку соответствия распределения случайной компоненты (остаточной последовательности) нормальному закону распределения можно произвести лишь приближенно с помощью исследования показателей асимметрии и эксцесса, так как временные ряды не очень велики.

## **1.2. Критерии оценки качества прогнозирования временных рядов**

С целью оценки качества прогноза, созданного с применением формализованных методов, общепринято применять подобные характеристики для надёжности, точности, достоверности, ошибки прогноза.

Под надёжностью прогнозных расчётов понимается мера неопределённости поведения объекта прогнозирования во времени.

Достоверность прогноза обуславливается вероятностью реализации прогноза для установленного варианта или доверительного интервала.

Точность прогноза определяет интервальный разброс прогнозных траекторий при фиксированном уровне достоверности [11].

Ошибки прогноза представляют собой меру отличия прогнозных оценок от реальных значений состояния прогнозируемого объекта.

Однако, изложить подобные свойства как надёжность, точность, достоверность, определить ошибки прогноза не представляется возможным, поскольку прогнозные результаты не с чем сравнивать. По этой причине перед разработчиками прогнозов встаёт проблема: «Как оценить качество прогноза ещё до его реализации?». Установленные шаги в сторону усовершенствования качества прогноза можно сделать, изучив факторы, влияющие на показатели качества прогноза.

Качество исходной информации, в свою очередь, обуславливается:



- точностью экономических измерений;
- качеством выборки;
- отсутствием погрешностей согласования (данные ошибки возникают в тех случаях, когда исходная информация для проведения прогнозных расчётов подготавливается различными специалистами, использующими разные методологические подходы) [10].

Погрешности, сопряженные с выбором модели прогноза, появляются вследствие упрощения, несовершенства теоретических построений или неадекватности моделей прогнозируемым социально-экономическим процессам. В некоторых случаях для прогнозирования процессов, проходящих в России, применяются модели, созданные иностранными экспертами и хорошо себя зарекомендовавшие для прогнозирования аналогичных процессов в других государствах. Но необходимо иметь в виду, что данные модели могут быть неадекватны по отношению к общественно-экономическим процессам, происходящим в нашей стране, и их применение может послужить причиной к значительным ошибкам и просчетам.

Итог прогноза, созданного формализованным методом, чаще всего проявляется количественным показателем, которому может быть дана точечная ( $\bar{y}_t$ ) и (или) интервальная оценка ( $\bar{y}_t$ ).

Точечная оценка – это единична оценка прогнозного параметра. Точечные значения экономических величин лишены содержания, так как имеют нулевую вероятность. Для устранения этого недостатка прогноз должен быть дан в виде интервала значений [10].

Интервальная оценка – это числовой интервал (доверительный интервал), в котором, вероятно, находится прогнозный параметр[10].

Достоверность прогноза тем выше, чем меньше величина ошибки, которая представляет собой разницу между прогнозируемыми и фактическими значениями исследуемой величины. Понятие точности

прогноза и способы ее оценки отличаются от точности исходных данных. Точность исходных данных может быть однозначно оценена на этапе сбора информации как степень приближения результатов измерений к истинному значению измеряемой величины. В практике зачастую количественную оценку точности заменяют указанием ошибки измерения (погрешности), которая определяется как разница между оцениваемым результатом и результатом, полученным более точным методом. В прогнозных значениях (до наступления прогнозируемого события) точность обычно также выражается как погрешность, но с помощью вероятностных пределов отклонения фактической величины от прогнозируемого значения, которые принято называть доверительным интервалом.

Отметим, то что прогнозируемые значения обязаны реализоваться в соответствующее время с указанной вероятностью и находится внутри некоторой доверительной области, ширина которой зависит от заданной вероятности.

Очевидно, что достоверность прогноза максимальна при построении точечного прогноза. Однако построить его со значительной степенью вероятности часто не представляется возможным.

Под достоверностью прогноза понимается вероятность осуществления прогноза в заданном доверительном интервале  $\pm\delta$  [8].

Существуют неформальный и формальный способы определения доверительного интервала. Неформально доверительный интервал может быть определен экспертами с учетом степени изменчивости фактических значений показателей вокруг расчетных (теоретических) значений в прошлом и возможности деформации в будущем. При этом экспертам может быть предложено оценить суммарную величину ошибки или степень влияния различных составляющих на точность прогноза. Суммарная ошибка решения прогнозной задачи определяется по формуле [6]:

$$\delta_{\zeta} = \delta_{\varepsilon} + \delta_i + \delta_{\lambda} + \delta_{\chi} + \delta_{\varsigma} \quad (1.3)$$

где  $\delta_t$  – суммарная ошибка;

$\delta_\varepsilon$  – ошибки информации, обусловленные неадекватностью описания объекта, погрешностями получения и обработки информации;

$\delta_i$  – ошибки метода прогнозирования, вызванные невозможностью идеального выбора метода для данного объекта, а также обязательной схематичностью метода;

$\delta_\lambda$  – ошибки вычислительных процедур;

$\delta_\chi$  – ошибки, допущенные человеком и обусловленные субъективными факторами (низкая квалификация, восторженность, пессимизм);

$\delta_\zeta$  – нерегулярная составляющая ошибки, обусловленная возможностью появления непредсказуемых изменений в объекте.

Формально границы доверительного интервала можно определить на основе оценки изменчивости уровней ряда. Чем выше эта изменчивость, тем менее точной может быть расчетная величина и тем шире должен быть доверительный интервал при одной и той же вероятности прогноза.

На практике, получая прогнозный результат в виде точечного значения  $Y_t$ , необходимо указать и возможную величину ошибки  $\delta$ , т.е. перейти к интервальному прогнозу по формуле

$$\hat{y}_t = \bar{y}_t \pm \delta \quad (1.4)$$

где  $\bar{y}_t$  – точечное значение прогнозной характеристики;

$\hat{y}_t$  – интервальное значение прогнозной характеристики;

$\delta$  – вероятная ошибка прогноза.

### 1.3. Ошибки прогнозирования

Важный аспект оценки качества прогнозов – определение причин ошибок в прогнозировании.

Остановимся на наиболее распространенных [24]:

1. Недоучет условности прогнозов. Перед прогнозированием ставится невыполнимая задача точных количественных оценок будущего, т.е. игнорируется вероятностный характер прогнозной информации. Избежать этой ошибки можно, если разрабатывать условные прогнозы.

2. Противоречие между требованиями определенной точности и достоверности прогноза, с одной стороны, и сложностью поставленной задачи и глубиной прогнозирования – с другой.

3. Неполнота, недостаточная разработанность теории, на которой разработан прогноз [24]. Качественное экономическое и социальное прогнозирование невозможно без теоретического осмысления сущности прогнозируемых процессов, выявления оценки логики их протекания, выявления взаимосвязей и взаимозависимостей. Эта причина существенна для прогнозирования многих процессов в российской экономике.

4. Сложность объекта прогнозирования. Трудно прогнозируемыми являются события нерегулярные, протекающие без определенных закономерностей.

5. Ошибки исходных данных, которые могут быть связаны с неточностью экономических измерений, нетипичностью выборки, искажением данных при их агрегировании.

Отметим также, что процесс оценки качества прогноза необходимо начинать с проверки качества исходной информации, корректности теоретических и методических положений, составляющих основу разрабатываемого прогноза. Качество исходной информации, точность измерений задают пределы использования математического аппарата для ее обработки, а также границы достижимой достоверности получаемых результатов.

## 2. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ПРОГНОЗИРОВАНИИ И ОЦЕНИВАНИИ ОШИБОК МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

### 2.1. Базовые модели временных рядов

#### 1. Регрессионные модели

Существует много задач, требующих изучения отношения между двумя и более переменными. Для решения таких задач используется регрессионный анализ [15]. В настоящее время регрессия получила широкое применение, включая задачи прогнозирования и управления. Целью регрессионного анализа является определение зависимости между исходной переменной и множеством внешних факторов (регрессоров). При этом коэффициенты регрессии могут определяться по методу наименьших квадратов или методу максимального правдоподобия [20].

Линейная регрессионная модель. Самым простым вариантом регрессионной модели является линейная регрессия. В основу модели положено предположение, что существует дискретный внешний фактор  $X(t)$ , оказывающий влияние на исследуемый процесс  $Z(t)$ , при этом связь между процессом и внешним фактором линейна. Модель прогнозирования на основании линейной регрессии описывается уравнением

$$Z(t) = \alpha_0 + \alpha_1 X(t) + \varepsilon_t, \quad (2.1)$$

где  $\alpha_0$  и  $\alpha_1$  – коэффициенты регрессии;  $\varepsilon_t$  – ошибка модели. Для получения прогнозных значений  $Z(t)$  в момент времени  $t$  необходимо иметь значение  $X(t)$  в тот же момент времени  $t$ , что редко выполнимо на практике.

Множественная регрессионная модель. На практике на процесс  $Z(t)$  оказывают влияние целый ряд дискретных внешних факторов  $X_1(t), \dots, X_s(t)$ . Тогда модель прогнозирования имеет вид

$$Z(t) = \alpha_0 + \alpha_1 X_1(t) + \alpha_2 X_2(t) + \dots + \alpha_s X_s(t) + \varepsilon_t \quad (2.2)$$

Недостатком данной модели является то, что для вычисления будущего значения процесса  $Z(t)$  необходимо знать будущие значения всех факторов  $X_1(t), \dots, X_s(t)$ , что почти невыполнимо на практике.

## 2. Авторегрессионные модели

В основу авторегрессионных моделей заложено предположение о том, что значение процесса  $Z(t)$  линейно зависит от некоторого количества предыдущих значений того же процесса  $Z(t-1), \dots, Z(t-p)$ .

Авторегрессионная модель скользящего среднего. В области анализа временных рядов модель авторегрессии (autoregressive, AR) и модель скользящего среднего (moving average, MA) является одной из наиболее используемых [5].

Модель авторегрессии является исключительно полезной для описания некоторых встречающихся на практике временных рядов. В этой модели текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и импульса, который называется «белым шумом»,

$$Z(t) = C + \phi_1 Z(t-1) + \phi_2 Z(t-2) \dots \phi_p Z(t-p) + \varepsilon_t \quad (2.3)$$

Формула (2.3) описывает процесс авторегрессии порядка  $p$ , который в литературе часто обозначается  $AR(p)$ , здесь  $C$  – вещественная константа,  $\phi_0, \dots, \phi_p$  – коэффициенты,  $\varepsilon_t$  – ошибка модели. Для определения  $\phi_i$  и  $C$  используют метод наименьших квадратов или метод максимального правдоподобия [20].

Другой тип модели имеет большое значение в описании временных рядов и часто используется совместно с авторегрессией называется моделью скользящего среднего порядка  $q$  и описывается уравнением

$$z(t) = \frac{1}{q} (z(t-1) + Z(t-2) + \dots + Z(t-q)) + \varepsilon_t \quad (2.4)$$

В литературе процесс (2.3) часто обозначается  $MA(q)$ ; здесь  $q$  – порядок скользящего среднего,  $\varepsilon_t$  – ошибка прогнозирования. Модель скользящего среднего является по сути дела фильтром низких частот. Нужно отметить, что существуют простые, взвешенные, кумулятивные, экспоненциальные модели скользящего среднего.

## 3. Модели экспоненциального сглаживания

Модели экспоненциального сглаживания разработаны в середине XX века и до сегодняшнего дня являются широко распространенными в силу их простоты и наглядности.

Модель экспоненциального сглаживания (exponential smoothing, ES) применяется для моделирования финансовых и экономических процессов [22]. В основу экспоненциального сглаживания заложена идея постоянного пересмотра прогнозных значений по мере поступления фактических. Модель ES присваивает экспоненциально убывающие веса наблюдениям по мере их старения. Таким образом, последние доступные наблюдения имеют большее влияние на прогнозное значение, чем старшие наблюдения.

Функция модели ES имеет вид:

$$\begin{aligned} Z(t) &= S(t) + \varepsilon_t \\ S(t) &= \alpha \cdot Z(t-1) + (1-\alpha) \cdot S(t-1) \end{aligned} \quad (2.5)$$

где  $\alpha$  – коэффициент сглаживания,  $0 < \alpha < 1$ ; начальные условия определяются как  $S(1) = Z(0)$ . В данной модели каждое последующее сглаженное значение  $S(t)$  является взвешенным средним между предыдущим значением временного ряда  $Z(t)$  и предыдущего сглаженного значения  $S(t-1)$ .

Модель Хольта или двойное экспоненциальное сглаживание применяется для моделирования процессов, имеющих тренд. В этом случае в модели необходимо рассматривать две составляющие: уровень и тренд [25]. Уровень и тренд сглаживаются отдельно:

$$\begin{aligned} Z(t) &= S(t) + \varepsilon_t \\ S(t) &= \alpha \cdot Z(t-1) + (1-\alpha) \cdot (S(t-1) - B(t-1)) \\ B(t) &= \gamma \cdot (S(t-1) - S(t-2)) + (1+\gamma) \cdot B(t-1) \end{aligned} \quad (2.6)$$

Здесь  $\alpha$  – коэффициент сглаживания уровня,  $\gamma$  – коэффициент сглаживания тренда.

#### 4. Нейросетевые модели

В настоящее время самой популярной среди структурных моделей является модель на основе искусственных нейронных сетей (artificial neural network, ANN). Нейронные сети состоят из нейронов (Рисунок 2.1) [7].

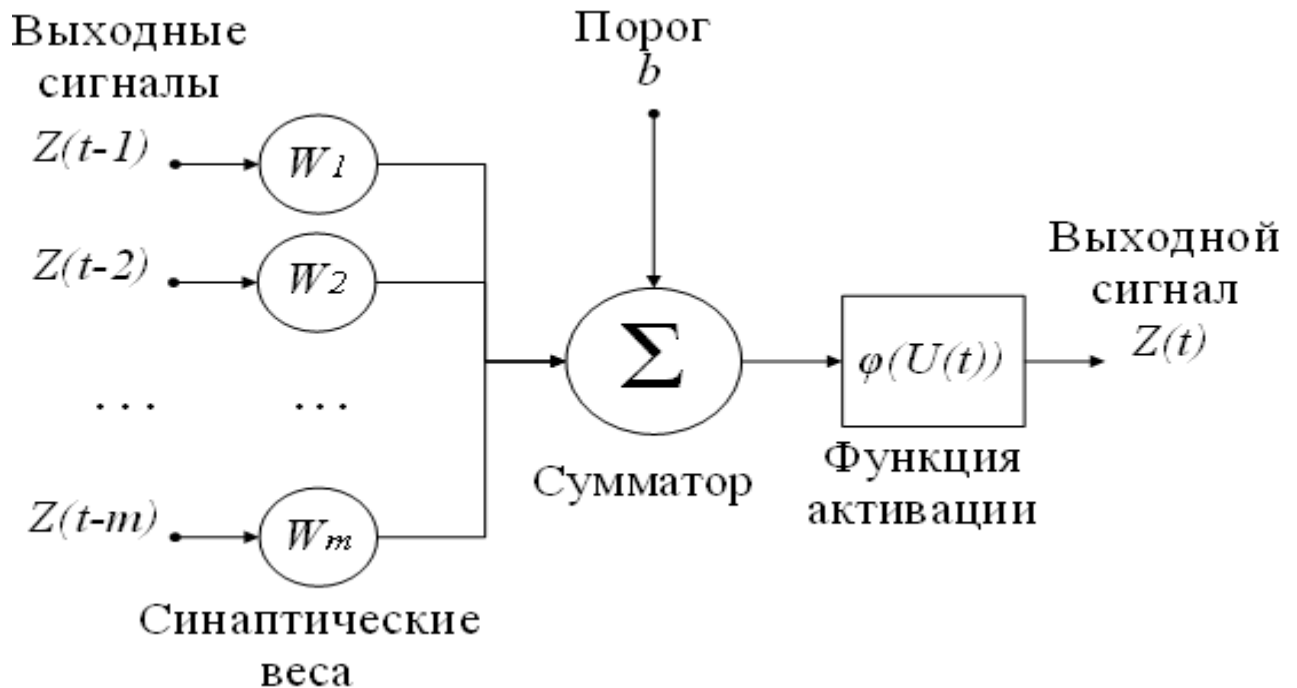


Рис. 2.1. – Нелинейная модель нейрона

Модель нейрона можно описать парой уравнений:

$$U(t) = \sum_{i=1}^m \omega_i \cdot Z(t-i) + b$$

$$Z(t) = \varphi(U(t)) \quad (2.7)$$

где  $Z(t-1), \dots, Z(t-i)$  – входные сигналы;  $\omega_1, \dots, \omega_m$  – синаптические веса нейрона;  $p$  – порог;  $\varphi(U(t))$  – функция активации.

Функция активации бывает трех основных типов [22]:

- функция единичного скачка;
- кусочно-линейная функция;
- сигмоидальная функция.

Способ связи нейронов определяет архитектуру нейронной сети. В зависимости от способа связи нейронов сети делятся на однослойные нейронные сети прямого распространения, многослойные нейронные сети прямого распространения, рекуррентные нейронные сети.



На рисунке 2.2 представлена структура трехслойной нейронной сети прямого распространения [19].

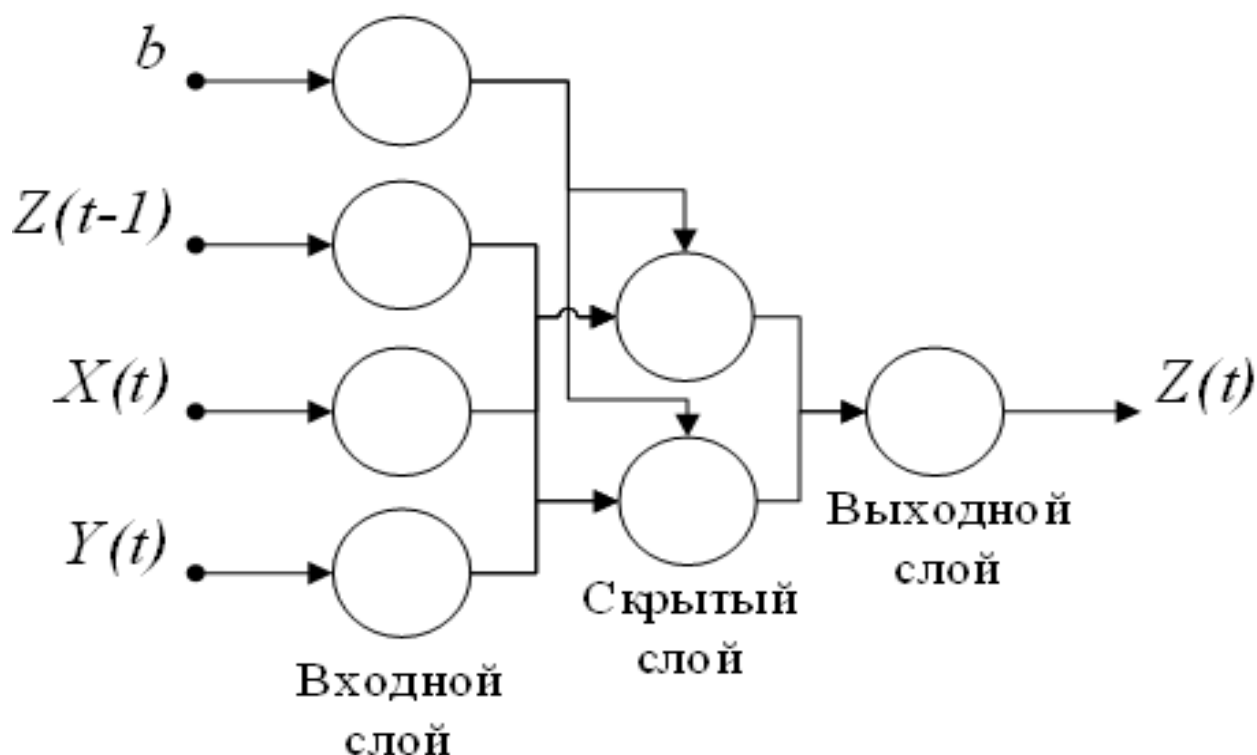


Рис. 2.2. – Трехслойная нейронная сеть прямого распространения

Таким образом, при помощи нейронных сетей возможно моделирование нелинейной зависимости будущего значения временного ряда от его фактических значений и от значений внешних факторов. Нелинейная зависимость определяется структурой сети и функцией активации.

## 5. Модели на базе цепей Маркова

Модели прогнозирования на основе цепей Маркова (Markov chain model) предполагают, что будущее состояние процесса зависит только от его текущего состояния и не зависит от предыдущих [13]. В связи с этим процессы, моделируемые цепями Маркова, должны относиться к процессам с короткой памятью.

Пример цепи Маркова для процесса, имеющего три состояния, представлен на рисунке 2.3 [19].

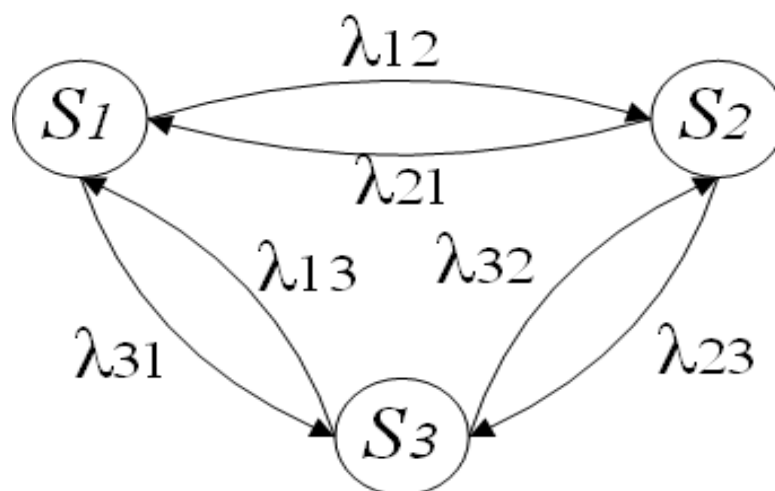


Рис. 2.3. – Цепь Маркова с тремя состояниями

Здесь  $S_1, \dots, S_3$  – состояния процесса  $Z(t)$ ;  $\lambda_{23}$  – вероятность перехода из состояния  $S_1$  в состояние  $S_2$ ,  $\lambda_{23}$  – вероятность перехода из состояния  $S_2$  в состояние  $S_3$  и т.д. При построении цепи Маркова определяется множество состояний и вероятности переходов. Есть текущее состояние процесса  $S_i$ , то в качестве будущего состояния процесса выбирается такое состояние  $S_j$ , вероятность перехода в которое (значение  $\lambda_{ij}$ ) максимальна.

Таким образом, структура цепи Маркова и вероятности перехода состояний определяют зависимость между будущим значением процесса и его текущим значением.

#### 6. Модели на базе классификационно-регрессионных деревьев

Классификационно-регрессионные деревья (classification and regression trees, CART) являются еще одной популярной структурной моделью прогнозирования временных рядов [17]. Структурные модели CART разработаны для моделирования процессов, на которые оказывают влияние как непрерывные внешние факторы, так и категориальные. Если внешние факторы, влияющие на процесс  $Z(t)$ , непрерывны, то используются регрессионные деревья; если факторы категориальные, то – классификационные деревья. В случае, если необходимо учитывать факторы обоих типов, то используются смешанные классификационно-регрессионные деревья [7].



Рис. 2.4. – Бинарное классификационно-регрессионное дерево

Согласно модели CART, прогнозное значение временного ряда зависит от предыдущих значений, а также некоторых независимых переменных. На приведенном на рисунке 2.4 [7] примере сначала предыдущее значение процесса сравнивается с константой  $Z_0$ . Если значение  $Z(t-1) < Z_0$ , то выполняется следующая проверка:  $X(t) > X_{11}$ . Если неравенство не выполняется, то  $Z(t) = C_3$ , иначе проверки продолжают до того момента, пока не будет найден лист дерева, в котором происходит определение будущего значения процесса  $Z(t)$ . Важно, что при определении значения в расчет принимаются как непрерывные переменные, например,  $X(t)$ , так и категориальные  $Y$ , для которых выполняется проверка присутствия значения в одном из заранее определенных подмножеств. Значения пороговых констант, например,  $Z_0, X_{11}$ , а также подмножеств  $Y_{11}, Y_{12}$  выполняется на этапе обучения дерева [17]. Таким образом, CART моделирует зависимость будущей величины процесса  $Z(t)$  при помощи структуры дерева, а также пороговых констант и подмножеств.

## 7. Другие модели и методы прогнозирования

Кроме классов моделей прогнозирования, рассмотренных выше, существуют менее распространенные модели и методы прогнозирования. Главным недостатком моделей и методов, упомянутых в настоящем разделе, является недостаточная методологическая база, т. е. недостаточно подробное описание возможностей как моделей, так и путей определения их параметров. Кроме того, в открытом доступе можно найти лишь небольшое количество статей, посвященных применению данных методов.

Метод опорных векторов (support vector machine, SVM) применяется, например, для прогнозирования движения рынков [22] и цен на электроэнергию [5]. В основу метода положена классификация, производимая за счет перевода исходных временных рядов, представленных в виде векторов, в пространство более высокой размерности и поиска разделяющей гиперплоскости с максимальным зазором в этом пространстве. Алгоритм SVM работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора [13]. При этом задача прогнозирования решается таким образом, что на этапе обучения классификатора выявляются независимые переменные (внешние факторы), будущие значения которых определяют в какой из определенных ранее подклассов попадет прогноз  $Z(t)$ .

Генетический алгоритм (genetic algorithm, GA) был разработан и часто применяется для решения задач оптимизации, а также поисковых задач. Однако некоторые модификации GA позволяют решать задачи прогнозирования.

Алгоритм прогнозирования на основе GA позволяет принимать в расчет более 15 внешних факторов, используя базовый GA [14]. Принцип работы основан на том, что исходные значения процесса  $Z(t)$  и внешних факторов  $X_1(t), \dots, X_S(t)$  раскладывают в наборы, состоящие из 0 и 1, которые называют генотипами. Далее применяют ряд преобразований: скрещивание и

мутирование для формирования преобразованных наборов, которые называются фенотипами. Исходные и полученные наборы исследуются с использованием функции приспособленности. Если решение получилось неудовлетворительным, то снова производится скрещивание и мутирование, в результате чего получается еще более новые наборы (новое поколение), которые снова оцениваются. Итеративный процесс продолжается до тех пор, пока решение не будет удовлетворительным.

Модель на основе передаточных функций (transfer function, TF) применяется для прогнозирования процесса  $Z(t)$  с учетом внешнего фактора  $X(t)$  [15]. Уравнение, отражающее зависимость будущего значения имеет вид

$$Z(t) = v(B)X(t) + \eta(t) \quad (2.8)$$

где  $B$  – оператор сдвига  $BZ(t) = Z(t-1), \dots, B_k Z(t) = Z(t-k)$ . Временной ряд  $\eta(t)$  характеризует внешнее возмущение. При этом функция  $v(B)$  имеет вид

$$v(B) = v_0 + v_1 B + v_2 B^k + \dots \quad (2.9)$$

Коэффициенты функции (2.9)  $v_i$  описывают динамические отношения между процессами  $Z(t)$  и  $X(t)$ .

## **2.2. Сравнительный анализ моделей прогнозирования. Достоинства и недостатки моделей**

Регрессионные модели и методы. К достоинствам данных моделей прогнозирования относят простоту, гибкость, а также единообразие их анализа и проектирования. При использовании линейных регрессионных моделей результат прогнозирования может быть получен быстрее, чем при использовании остальных моделей. Кроме того, достоинством является прозрачность моделирования, т.е. доступность для анализа всех промежуточных вычислений.

Основным недостатком нелинейных регрессионных моделей является сложность определения вида функциональной зависимости, а также трудоемкость определение параметров модели. Недостатками линейных

регрессионных моделей являются низкая адаптивность и отсутствие способности моделирования нелинейных процессов [18].

Авторегрессионные модели и методы. Важными достоинствами данного класса моделей являются их простота и прозрачность моделирования. Еще одним достоинством является единообразие анализа и проектирования. На сегодняшний день данный класс моделей является одним из наиболее популярных, а потому в открытом доступе легко найти примеры применения авторегрессионных моделей для решения задач прогнозирования временных рядов различных предметных областей [4].

Недостатками данного класса моделей являются: большое число параметров модели, идентификация которых неоднозначна и ресурсоемка; низкая адаптивность моделей, а также линейность и, как следствие, отсутствие способности моделирования нелинейных процессов, часто встречающихся на практике [23].

Модели и методы экспоненциального сглаживания. Достоинствами данного класса моделей являются простота и единообразие их анализа и проектирования. Данный класс моделей чаще других используется для долгосрочного прогнозирования [16].

Недостатком данного класса моделей прогнозирования является отсутствие гибкости.

Нейросетевые модели и методы. Основным достоинством нейросетевых моделей является нелинейность, т.е. способность устанавливать нелинейные зависимости между будущими и фактическими значениями процессов. Другими важными достоинствами являются: адаптивность, масштабируемость (параллельная структура ANN ускоряет вычисления) и единообразие их анализа и проектирования [15].

При этом недостатками ANN являются отсутствие прозрачности моделирования; сложность выбора архитектуры, высокие требования к непротиворечивости обучающей выборки; сложность выбора алгоритма обучения и ресурсоемкость процесса их обучения [21].

Модели и методы на базе цепей Маркова. Простота и единообразие анализа и проектирования являются достоинствами моделей на базе цепей Маркова.

Недостатком данных моделей является отсутствие возможности моделирования процессов с длинной памятью [23].

Модели на базе классификационно-регрессионных деревьев. Достоинствами данного класса моделей являются: масштабируемость, за счет которой возможна быстрая обработка сверхбольших объемов данных; быстрота и однозначность процесса обучения дерева (в отличие от ANN) [13], а также возможность использовать категориальные внешние факторы.

Недостатками данных моделей являются неоднозначность алгоритма построения структуры дерева; сложность вопроса останова т.е. вопроса о том, когда стоит прекратить дальнейшие ветвления; отсутствие единообразия их анализа и проектирования [16].

Достоинства и недостатки моделей и методов систематизированы в таблице 2.1.

Таблица 2.1.

Сравнение моделей и методов прогнозирования

Модель и метод	Достоинства	Недостатки
Регрессионные модели и методы	простота, гибкость, прозрачность моделирования; единообразие анализа и проектирования	сложность определения функциональной зависимости; трудоемкость нахождения коэффициентов зависимости; отсутствие возможности моделирования нелинейных процессов

Продолжение таблицы 2.1.

Модель и метод	Достоинства	Недостатки
Авторегрессионные модели и методы	простота, прозрачность моделирования; единообразие анализа и проектирования; множество примеров применения	трудоемкость и ресурсоемкость идентификации моделей; невозможность моделирования нелинейностей; низкая адаптивность
Модели и методы экспоненциального сглаживания	простота моделирования; единообразие анализа и проектирования	недостаточная гибкость; узкая применимость моделей
Нейросетевые модели и методы	нелинейность моделей; масштабируемость, высокая адаптивность; единообразие анализа и проектирования; множество примеров применения	отсутствие прозрачности; сложность выбора архитектуры; жесткие требования к обучающей выборке; сложность выбора алгоритма обучения; ресурсоемкость процесса обучения
Модели и методы на базе цепей Маркова	простота моделирования; единообразие анализа и проектирования	невозможность моделирования процессов с длинной памятью; узкая применимость моделей
Модели и методы на базе классификационно-регрессионных деревьев	масштабируемость; быстрота и простота процесса обучения; возможность учитывать категориальные переменные	неоднозначность алгоритма построения дерева; сложность вопроса останова

Необходимо дополнительно отметить, что ни для одной из рассмотренных групп моделей (и методов) в достоинствах не указана точность прогнозирования. Это сделано в связи с тем, что точность прогнозирования того или иного процесса зависит не только от модели, но и от опыта исследователя, от доступности данных, от располагаемой аппаратной мощности и многих других факторов. Точность прогнозирования будет оцениваться для конкретных задач, решаемых в рамках данной работы.



На сегодняшний день наиболее распространенными моделями прогнозирования являются авторегрессионные модели (ARIMAX), а также нейросетевые модели (ANN). В статье [3], в частности, утверждается: «Without a doubt ARIMA(X) and GRACH modeling methodologies are the most popular methodologies for forecasting time series. Neural networks are now the biggest challengers to conventional time series forecasting methods». (Без сомнений модели ARIMA(X) и GARCH являются самыми популярными для прогнозирования временных рядов. В настоящее время главную конкуренцию данным моделям составляют модели на основе ANN.)

### **2.3. Информационные технологии и инструментарий оценивания прогнозирования**

В настоящее время любому предприятию для успешного стратегического и оперативного планирования не обойтись без детального анализа данных и прогнозирования различных экономических показателей. При воздействии сразу множества факторов очень трудно следить за всеми процессами одновременно с точки зрения контроля, принятия решений и управления.

Использование компьютерных технологий позволяет создать уникальную среду, в которой статистическая обработка данных становится не рутинным занятием, а увлекательным исследованием, позволяющим получать многовариантные решения. Рассмотрим применение современных пакетов прикладных программ при обработке экономических временных рядов.

В моделях, базирующихся на статистической информации, используют следующие основные подходы:

- вероятностный, с предположением о нормальности распределения изучаемых величин (модели математической статистики);
- геометрический, согласно которому данные не имеют

вероятностной природы и образуют в многомерном пространстве структуры с определенными свойствами;

- содержательный, предполагающий достижение определенных целей моделирования.

Первые два подхода реализуются в прикладной статистике, третий - при интеллектуальном анализе данных. В отличие от классических моделей статистики и эконометрики, при интеллектуальном анализе вид изучаемых связей не задается априори, а подбирается итеративно с целью наилучшего описания данных [12].

Стандартные статистические методы обработки данных включены в состав электронных таблиц (Excel, Lotus 1-2-3, Quattro Pro), математические пакеты общего назначения (Mathcad, MATLAB), а также в некоторые базы данных. Большими возможностями анализа и прогнозирования временных рядов обладают статистические пакеты, основную часть которых составляют специализированные пакеты и пакеты общего назначения.

Статистические пакеты общего назначения включают широкий диапазон статистических методов и используются на начальных этапах обработки при подборе статистической модели или метода анализа данных.

Это пакеты R-Studio, SAS, STATISTICA, MATLAB, SciPy, SPSS, MS Excel, и др. К этой группе можно отнести и эконометрические пакеты Stata.

Наилучший выбор статистического пакета для анализа данных зависит от характера решаемых задач, объема и специфики обрабатываемых данных, квалификации пользователей, имеющегося оборудования и т.д.

На рисунке 2.5 представлена классификация программного обеспечения ориентированного на обработку временных рядов по способу решения поставленных задач:

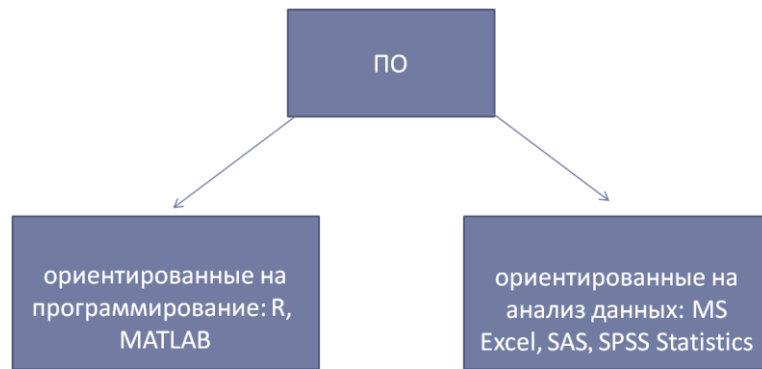


Рис. 2.5. – Классификация программного обеспечения для анализа и прогнозирования временных рядов

В таблице 2.2. представлена сравнительная характеристика информационных технологий, в которой выделены основные достоинства и недостатки технологий перечисленных выше.

Таблица 2.2

Сравнительная характеристика информационных технология анализа и прогнозирования временных рядов

Программный продукт	+	–
R-studio	<ul style="list-style-type: none"> <li>гибкость;</li> <li>наличие встроенных мощных аналитических средств;</li> <li>в части анализа временных рядов R имеет встроенную поддержку ARIMA;</li> <li>открытый код.</li> </ul>	<ul style="list-style-type: none"> <li>ориентация языка программирования на конкретную предметную область (финансовый сектор);</li> <li>сложность обучения;</li> </ul>
MATLAB	<ul style="list-style-type: none"> <li>поддержка работы с матрицами;</li> <li>удобный графический интерфейс;</li> <li>простота в работа.</li> </ul>	<ul style="list-style-type: none"> <li>дороговизна лицензии;</li> <li>неполная поддержка статистических функций.</li> </ul>
ScyPy	<ul style="list-style-type: none"> <li>широкие возможности по интеграции языка Python;</li> <li>высокая производительность математических операций;</li> <li>простота освоения и дальнейшего использования.</li> </ul>	<ul style="list-style-type: none"> <li>недоработанность продукта;</li> <li>узкое применение математического инструментария.</li> </ul>
MS Excel	<ul style="list-style-type: none"> <li>удобный интерфейс;</li> <li>простота освоения и дальнейшего использования.</li> </ul>	<ul style="list-style-type: none"> <li>ограниченный набор функций для анализа данных.</li> </ul>

Программный продукт	+	-
SAS	<ul style="list-style-type: none"> <li>• Гибкий интерфейс обмена данными;</li> <li>• наличие инструментария для работы с кластерами;</li> <li>• быстрота расчетов на громадных массивах данных.</li> </ul>	<ul style="list-style-type: none"> <li>• дороговизна лицензий;</li> <li>• сложность освоения.</li> </ul>
SPSS Statistics	<ul style="list-style-type: none"> <li>• удобный графический интерфейс.</li> </ul>	<ul style="list-style-type: none"> <li>• дороговизна лицензий;</li> <li>• отсутствие гибкости в расчетах.</li> <li>• Узкая ориентация (социальный сектор).</li> </ul>
Stats	<ul style="list-style-type: none"> <li>• удобный графический интерфейс;</li> <li>• Более низкая стоимость лицензий по сравнению с аналогичными пакетами</li> </ul>	<ul style="list-style-type: none"> <li>• узкая специализация.</li> </ul>

#### 2.4. Систематизация ошибок моделей прогнозирования и методов их расчета

Характеристики качества информационной пригодности моделей прогнозирования описывают, насколько достоверно, выбранная в качестве генератора прогноза, модель описывает ретроспективу исследуемого явления. Чем точнее построенная модель объясняла прошлое, тем больше вероятность того, что она будет удачно предсказывать будущее. Надежность моделей прогнозирования оценивается путем сравнения фактических и предсказанных значений. Эта разница позволяет проверить, применима ли к конкретным данным рассматриваемая модель и те предположения, на которых она основана. Основными оценочными характеристиками качества прогнозной модели являются нижеследующие показатели:

##### 1. Абсолютные ошибки прогнозирования

Первая группа базируется на расчете абсолютных ошибок и включает в себя оценки, основанные на вычислении величины  $e_t$ :

$$e_t = (y_t - f_t^{(m)}), \quad (2.11)$$

где  $y_t$  – измеренное значение в момент времени  $t$ ,  $f_t^{(0)}$  – прогнозное значение в момент времени  $t$

Средняя абсолютная ошибка (Mean Absolute Error, MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| = \underset{i=1, n}{mean} |e_i|, \quad (2.12)$$

где  $n$  – размер горизонта прогнозирования (forecast horizon),  $mean(\bullet)$  – операция нахождения среднего.

Медиана абсолютной ошибки (Median Absolute Error, MdAE) рассчитывается согласно формуле:

$$MdAE = \underset{i=1, n}{median} |e_i| \quad (2.13)$$

где  $median(\bullet)$  – операция нахождения медианы.

Квадрат средней квадратичной ошибки (Mean Square Error, MSE) рассчитывается по формуле

$$MSE = \frac{1}{n} \sum_{i=1}^n (e_i^2) = \underset{i=1, n}{mean}(e_i^2), \quad (2.14)$$

Средняя квадратичная ошибка (Root Mean Square Error, RMSE) рассчитывается согласно формуле:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i^2)} = \sqrt{\underset{i=1, n}{mean}(e_i^2)} \quad (2.15)$$

## 2. Процентные ошибки прогнозирования

Процентные ошибки рассчитываются на основании значения  $p_t$ :

$$p = \frac{|e_t|}{y_t} \quad (2.16)$$

Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error, MAPE):

$$MAPE = \frac{1}{n} \sum_{i=1}^n 100 \cdot |p_i| = \underset{i=1, n}{mean}(100 \cdot |p_i|) \quad (2.17)$$

Медиана абсолютной процентной ошибки (Median Absolute Percentage Error, MdAPE):

$$MdAPE = \underset{i = \overline{1, n}}{median} (100 \cdot |p_i|) \quad (2.18)$$

Средняя квадратичная процентная ошибка (Root Mean Square Percentage Error, RMSPE) вычисляется согласно:

$$RMSPE = \sqrt{\underset{i = \overline{1, n}}{maen} (100 \cdot |p_i|)} \quad (2.19)$$

Медиана квадратичной процентной ошибки:

$$RMdSPE = \sqrt{\underset{i = \overline{1, n}}{median} (100 \cdot |p_i|)^2} \quad (2.20)$$

### 3. Симметричные ошибки прогнозирования

Критерии, входящие в данную группу, рассчитываются на основании значения  $s_t$ , вычисляемого по формуле:

$$s_t = \frac{|e_t|}{y_t + f_t} \quad (2.21)$$

В группу входят следующие оценки.

Симметричная средняя абсолютная процентная ошибка:

$$sMAPE = \frac{1}{n} \sum_{i=1}^n 200 \cdot |s_i| = \underset{i = \overline{1, n}}{maen} (200 \cdot |s_i|) \quad (2.22)$$

Медиана средней абсолютной процентной ошибки:

$$sMdAPE = \underset{i = \overline{1, n}}{median} 200 \cdot |s_i| \quad (2.23)$$

Во избежание проблем связанных с делением на ноль была предложена модифицированная sMAPE – Modified sMAPE, msMAPE, в знаменателе которой появился дополнительный член:

$$msMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - f_i|}{\frac{|y_i| + |f_i|}{2} + S_i},$$

$$\text{где } S_i = \frac{1}{i-1} \sum_{k=1}^{i-1} |y_k - \bar{y}_{i-1}|, \bar{y}_{i-1} = \frac{1}{i-1} \sum_{k=1}^{i-1} y_k \quad (2.24)$$

### 4. Ошибки, основанные на оценке прогноза базовых модлей

Основой для расчета ошибок данной группы является величина  $r_t$ , определяемая следующим образом:

$$r_t = \frac{|e_t|}{(y_t - f_t^*)}, \quad (2.25)$$

где  $O$  – прогнозное значение, полученное с использованием базовой модели прогнозирования (benchmark model). Зачастую в качестве такой модели выступает наивная модель, для которой

$$f_t^* = y_{t-l}, \quad (2.26)$$

где  $l$  – величина лага (обычно  $l=1$ )

В данную группу входят следующие оценки. Средняя относительная абсолютная ошибка (Mean Relative Absolute Error, MRAE) рассчитывается по формуле:

$$MRAE = \frac{maen}{i = \overline{1, n}} |r_i|, \quad (2.27)$$

Медиана относительной абсолютной ошибки (Median Relative Absolute Error, MdRAE), вычисляемая согласно:

$$MdRAE = \frac{median}{i = \overline{1, n}} |r_i| \quad (2.28)$$

## 5. Масштабированные ошибки прогнозирования

В качестве базиса для расчета масштабированных ошибок используется величина  $qt$

$$q_t = \frac{|e_t|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - y_{i-1}|}. \quad (2.29)$$

В данную группу входит средняя абсолютная масштабированная ошибка (Mean Absolute Scaled Error, MASE) [2]:

$$MASE = \frac{maen}{i = \overline{1, n}} |q_i|, \quad (2.30)$$

Медиана абсолютной масштабированной ошибки находится следующим образом:

$$MdASE = \frac{median}{i = \overline{1, n}} |q_i| \quad (2.31)$$

Среднеквадратическая масштабированная ошибка (Root Mean Square Scaled Error, RMSSE), вычисляемая по формуле [3]:

$$RMSSE = \sqrt{\frac{maen}{i=1,n} (|q_i|)^2} \quad (2.32)$$

#### 6. Относительные оценки

В данную группу входят оценки, рассчитываемые на основе отношения ошибок, рассмотренных выше для анализируемой модели прогнозирования и базовой модели.

Относительная средняя абсолютная ошибка:

$$RMAE = \frac{MAE}{MAE^*} \quad (2.33)$$

где  $MAE$  и  $MAE^*$  – средние абсолютные ошибки для анализируемой модели прогнозирования и базовой модели соответственно, вычисляемые по формуле (2.12).

Относительное среднеквадратичное отклонение (Relative Root Mean Square Error, RelRMSE) вычисляется аналогично (2.33), только в правой части вычисляются оценки по формуле (2.15):

$$RRMSE = \frac{RMSE}{RMSE^*} \quad (2.34)$$

Оценка логарифм относительного среднеквадратичного отклонения (Log mean squared error ratio, LMR):

$$LMR = \log\left(\frac{RMSE}{RMSE^*}\right) \quad (2.35)$$

Следующая группа оценок рассчитывает число случаев, когда ошибка исследуемой модели прогнозирования превышает значение ошибки базовой модели. К таким оценкам относятся  $PB(MAE)$  – Percentage Better (MAE), вычисляемая по формуле:

$$PB(MAE) = 100\% \cdot \text{mean}(I\{MAE < MAE^*\}), \quad (2.36)$$

где  $I$  – оператор, принимающий значение ноль или единица в соответствии с выражением:

$$I(MAE) = \begin{cases} 0, & \text{если } MAE < MAE^*; \\ 1, & \text{в противном случае.} \end{cases} \quad (2.37)$$



## 7. Другие группы оценок

В данной группе собраны оценки, предложенные в различных исследованиях для устранения недостатков существующих и распространенных оценок.

Так для нивелирования недостатка связанного с зависимостью от шкалы предложена нормализованная среднеквадратическая ошибка (Normalized Root Mean Square Error nRMSE), вычисляемая по формуле:

$$nRMSE = \frac{1}{\bar{y}} \sqrt{\frac{maen}{i=1, n} (e_i^2)} \quad (2.38)$$

Для снижения влияния выбросов были предложены интегральные оценки Integral Normalized Mean Square Error, вычисляемые по формуле [1]:

$$inRSE = \frac{1}{\sum_{i=1}^n y_i} \sqrt{\frac{maen}{i=1, n} (e_i^2)} \quad (2.39)$$

В некоторых источниках оценка nRMSE рассчитывается следующим образом:

$$inRSE = \sqrt{\frac{\sum_{i=1}^n (e_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (2.40)$$

где  $\bar{y} = \frac{\text{Ошибка! За.}}{\dots}$

Другие нормализованные оценки получили название std\_APE, std\_MAPE и рассчитываются по формулам:

$$Std\_AE = \sqrt{\frac{\sum_{i=1}^n (e_i - MAE)^2}{n-1}} \quad (2.41)$$

$$Std\_APE = \sqrt{\frac{\sum_{i=1}^n (p_i - MAE)^2}{n-1}} \quad (2.42)$$

### **3. ОЦЕНКИ ОШИБОК ПРОГНОЗИРОВАНИЯ ПО МОДЕЛИ ПРОГНОЗИРОВАНИЯ ЧИСЛЕННОСТЬ НЕТРУДОУСТРОЕННЫХ ГРАЖДАН**

#### **3.1. Разработка модели прогнозирования с применением информационных технологий пакета MS EXCEL**

Одной из важнейших проблем современной экономики России является проблема безработицы. Проблема занятости и безработицы в наши дни становится неотъемлемым элементом жизни России, оказывающим существенное влияние не только на социально-экономическую, но и на политическую ситуацию в стране. Также безработица неразрывно связана с происходящими структурными сдвигами в национальной экономике и мировом хозяйстве. Она как сложное и противоречивое макроэкономическое явление остается в центре внимания и населения, и научной общественности.

Потеря работы для большинства людей означает неустойчивость экономического положения и снижение жизненного уровня. Занятость раскрывает один из важнейших аспектов социального развития человека, связанный с удовлетворением его потребностей в сфере труда. Как социально-экономическая категория занятости характеризует деятельность граждан, связанную с удовлетворением личных и общественных потребностей.

Для анализа временного ряда, нами были взяты данные о численности нетрудоустроенных граждан, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года.

Данный временной промежуток охватывает 111 месяцев. Данные представлены в таблице 3.1.

Таблица 3.1

Данные о численности нетрудоустроенных граждан, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года

Месяц	Годы									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Январь	67,8	54,7	31,3	46,4	37,9	32,9	26,4	25,2	26,1	22,6
Февраль	67,9	61,2	61	47	39,2	34,1	27,4	27	27,3	23,2
Март	66,7	66,8	61	46,3	38,1	33,1	26,6	26,5	27	23
Апрель	62,1	66,4	57,7	43,3	35,1	30	24,2	26,1	25	
Май	57,9	63,1	54,3	40	33	28	23	25	23,3	
Июнь	53,7	59,5	51	36,6	30,1	25,7	21,7	23,1	21,9	
Июль	51,9	58,3	47,7	34,3	28,6	24,5	21,4	22,6	20,8	
Август	47,3	57,3	44,8	32,9	27,2	23	20,4	21,8	20	
Сентябрь	43,6	55,4	42	31,1	25,9	20,7	18,6	20,4	18,1	
Октябрь	40,4	54,8	39,5	30,3	26,1	19,9	18,8	20,5	18	
Ноябрь	43,1	57,9	41,3	33,2	29,4	23,2	21,6	23,8	20,6	
Декабрь	47	58,9	43,7	34,5	30,3	24,6	22,6	23,9	21,4	

Пронумеруем в порядке возрастания и будем строить прогноз на 112-118 месяцы.

По оси  $O_x$  будут отображены месяцы по годам, по оси  $O_y$  тыс. человек нетрудоустроенных граждан, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года.

С помощью мастера диаграмм построим диаграмму временного ряда и на этой диаграмме отобразим линию тренда на рисунке 3.1.

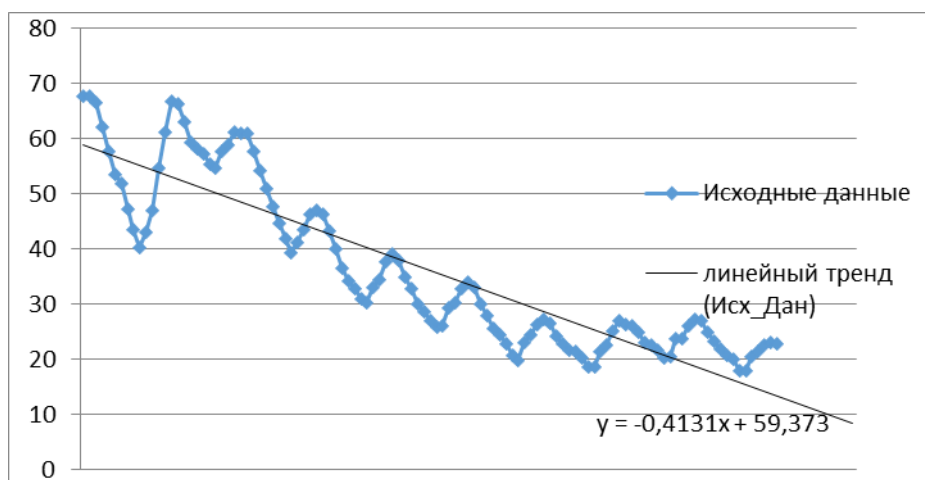


Рисунок 3.1 – Диаграмма количества нетрудоустроенных граждан в 2008-2017 годы

Анализируя диаграмму, представленную на рисунке 3.1. можно утверждать, что:

- имеется тенденция к уменьшению числа нетрудоустроенных граждан (линейный тренд с отрицательным углом наклона), обусловленная положительными изменениями социальной ситуации.
- - имеются сезонные колебания с максимумом в зимние периоды и минимумом в летние периоды.

Предварительный графический анализ ряда задает направление дальнейшего исследования рисунок 3.2:

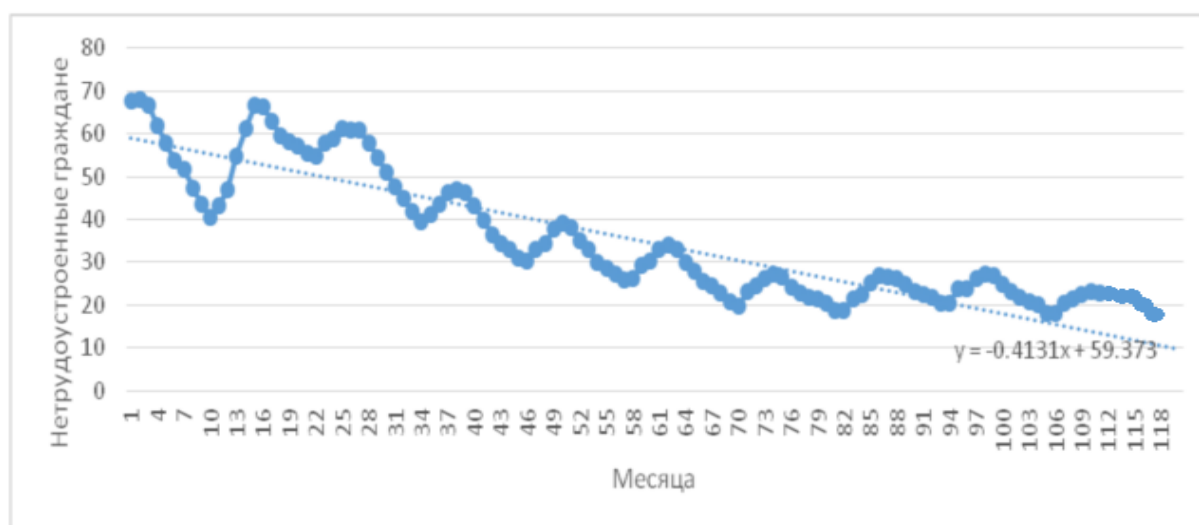


Рис. 3.2. –Прогноз нетрудоустроенных граждан на январь-март 2017 года

Уравнение линии тренда  $y = -0,4131x + 59,373$

В таблице 3.2 представлены значения по месяцам линии тренда, а также прогнозные значения на 2017 год.

Таблица 3.2

Значения линии тренда по месяцам

месяцы	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Январь	58,96	54,00	49,05	44,09	39,13	34,17	29,22	24,26	19,30	14,35
Февраль	58,55	53,59	48,63	43,68	38,72	33,76	28,80	23,85	18,89	13,93
Март	58,13	53,18	48,22	43,26	38,30	33,35	28,39	23,43	18,48	13,52
Апрель	57,72	52,76	47,81	42,85	37,89	32,93	27,98	23,02	18,06	13,11
Май	57,31	52,35	47,39	42,44	37,48	32,52	27,56	22,61	17,65	12,69
Июнь	56,89	51,94	46,98	42,02	37,07	32,11	27,15	22,19	17,24	12,28
Июль	56,48	51,52	46,57	41,61	36,65	31,70	26,74	21,78	16,82	11,87
Август	56,07	51,11	46,15	41,20	36,24	31,28	26,33	21,37	16,41	11,45
Сентябрь	55,66	50,70	45,74	40,78	35,83	30,87	25,91	20,95	16,00	11,04
Октябрь	55,24	50,28	45,33	40,37	35,41	30,46	25,50	20,54	15,58	10,63
Ноябрь	54,83	49,87	44,91	39,96	35,00	30,04	25,09	20,13	15,17	10,21
Декабрь	54,42	49,46	44,50	39,54	34,59	29,63	24,67	19,72	14,76	9,80

Представленная модель тренда описывает общую тенденцию поведения ряда. Сделать на базе этой модели достаточно точный прогноз ежемесячного числа нетрудоустроенных граждан на 2017 год нельзя. Так как большая сезонность (годовая) изменчивости ряда. Для построения месячного прогноза необходимо оценить сезонные эффекты или сезонные индексы.

### 3.2. Динамика и прогнозирование временного ряда численность нетрудоустроенных граждан Алтайского края

На рисунке 3.2 видно, что величина сезонных колебаний пропорциональна среднему уровню нетрудоустроенных граждан. Для этого используем описание сезонных колебаний с применением смешанной мультипликативной модели:  $x_t = tr_t \cdot S_t + \varepsilon_t$ . Задача состоит в оценке

значений  $S_t$ , в предположении, что  $S_{t/p} = S_t$ , где  $p$  – период последовательности.

В рассматриваемом примере период последовательности  $p$  составляет 12 ( $p=12$ ), а всего периодов 9.

Для получения сезонных индексов рассчитаем отношение исходных данных на тренд.

Полученные значения представлены в таблице 3.3.

Полученные сезонные индексы позволят провести удаление сезонных эффектов, при этом исходные данные ряда разделим на соответствующие сезонные индексы. Данная процедура носит название сезонного выравнивания или сезонной коррекцией ряда. Результаты сезонного выравнивания приведены на диаграмме на рисунке 3.3.

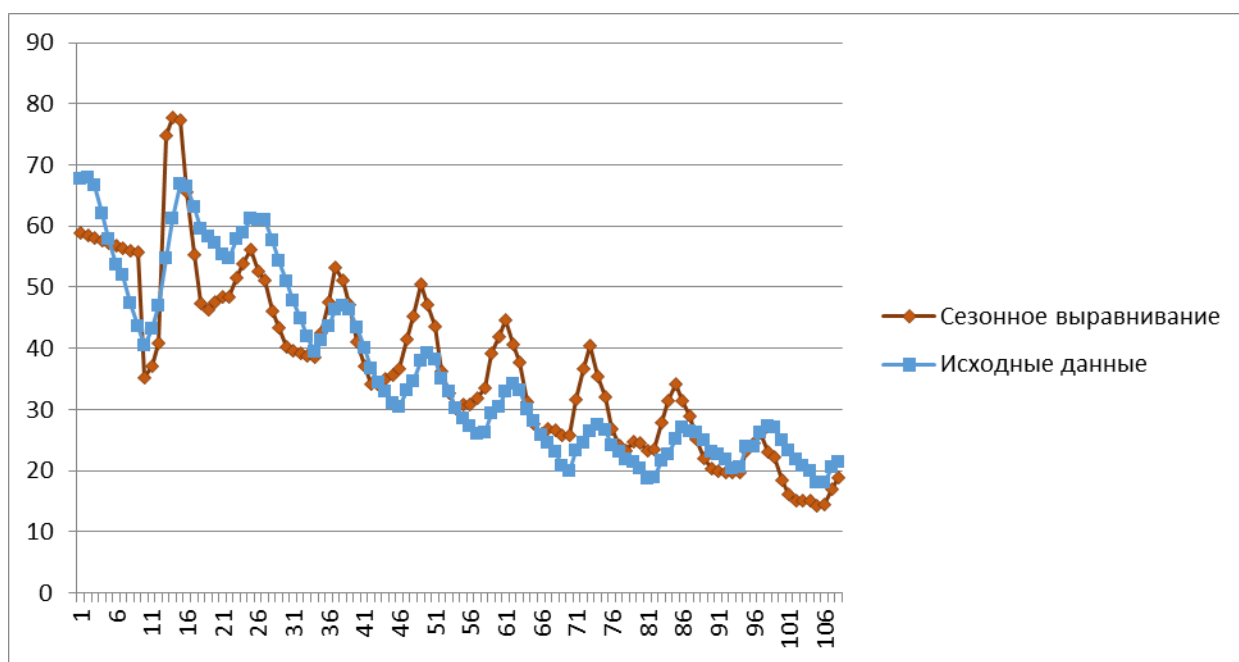


Рис.3.3 – Сезонное выравнивание ряда

Как видно из диаграммы, выровненный ряд имеет выраженную тенденцию убывания.

Таблица 3.3

## Сезонные индексы числа нетрудоустроенных граждан

Месяц	Годы									
	2008	2009	2010	2011	2012	2013	2014	2015	2016	Сезонный индекс
Январь	1,149934	0,731328	1,089793	0,871434	0,75055	0,737013	0,653402	0,73729	0,997975	0,85763544
Февраль	1,159756	0,786082	1,160979	0,919525	0,830887	0,839998	0,772229	0,861048	1,182403	0,94587856
Март	1,147355	0,86372	1,190895	0,981992	0,872441	0,876052	0,830245	0,915996	1,21225	0,98788289
Апрель	1,075872	1,012912	1,24986	1,052433	0,968539	0,962723	0,903593	1,038768	1,35217	1,06854111
Май	1,010339	1,142013	1,254308	1,076126	1,012449	1,010047	0,95127	1,132246	1,44527	1,11489644
Июнь	0,943854	1,256194	1,265054	1,070221	0,994651	0,992572	0,936933	1,130869	1,461347	1,116855
Июль	0,918888	1,258448	1,206956	1,010525	0,926322	0,910896	0,864984	1,133787	1,384045	1,06831678
Август	0,843615	1,205342	1,145736	0,942598	0,8805	0,860969	0,834413	1,105847	1,320121	1,01546011
Сентябрь	0,783396	1,145614	1,085568	0,870956	0,812074	0,800414	0,799228	1,040822	1,270537	0,95651211
Октябрь	1,149934	1,131509	1,024333	0,824327	0,780301	0,772985	0,800356	1,037606	1,236351	0,973078
Ноябрь	1,159756	1,121089	0,970668	0,79861	0,750564	0,735242	0,774929	1,020227	1,218724	0,94997878
Декабрь	1,147355	1,092747	0,918219	0,762563	0,722933	0,670573	0,717817	0,973529	1,131427	0,90412922

## Прогнозные значения временного ряда

месяцы	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Январь	58,95	74,79	56,24	53,24	50,50	44,64	40,40	34,18	26,15	26,35
Февраль	58,54	77,85	52,54	51,11	47,18	40,60	35,48	31,36	23,09	24,53
Март	58,13	77,33	51,22	47,14	43,67	37,78	32,04	28,93	22,27	23,28
Апрель	57,72	65,55	46,16	41,14	36,24	31,16	26,78	25,13	18,49	19,11
Май	57,30	55,25	43,29	37,17	32,59	27,72	24,18	22,08	16,12	17,69
Июнь	56,89	47,3653	40,31	34,19	30,26	25,89	23,16	20,43	14,99	14,12
Июль	56,48	46,32	39,52	33,94	30,87	26,90	24,74	19,93	15,03	13,32
Август	56,06	47,53	39,10	34,90	30,89	26,71	24,45	19,71	15,15	13,45
Сентябрь	55,65	48,35	38,68	35,70	31,89	25,86	23,27	19,60	14,25	12,41
Октябрь	35,13	48,43	38,56	36,75	33,45	25,74	23,49	19,76	14,56	13,14
Ноябрь	37,16	51,64	42,54	41,57	39,17	31,55	27,87	23,33	16,90	15,23
Декабрь	40,96	53,90	47,59	45,24	41,91	36,69	31,48	24,55	18,91	17,98

### 3.3. Расчеты оценок ошибок прогнозирования

Основываясь на подраздел 2.3 данной работы, произведем оценку ошибок прогнозирования временного ряда проанализированного в пунктах 3.1 и 3.2. основанного на данных о численности нетрудоустроенных граждан Алтайского края.

Для оценки нами был выбран временной промежуток июнь 2016 – март 2017. Месяцам присвоены числовые значения 102–111 соответственно. Измеренные значения обозначим  $y_{102}-y_{111}$ . Прогнозные значения, полученные в результате исследования временного ряда обозначим  $f_{102}-f_{111}$ . Соотношение прогнозных и реальных данных показаны на рисунке 3.4



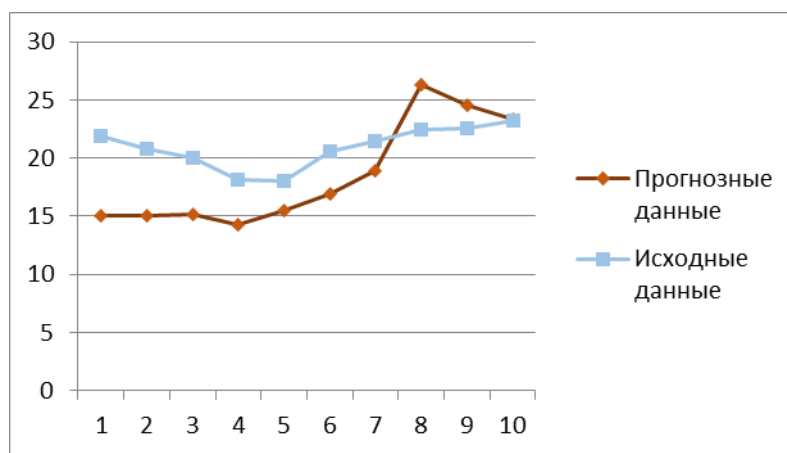


Рис. 3.4. – Исходные и прогнозные данные период с июня 2016-март 2017

### 1. Абсолютные ошибки прогнозирования

Высчитаем значение  $e_{102} - e_{111}$  по формуле 2.11:

$$e_{102} = 2,25;$$

$$e_{103} = 1,79;$$

$$e_{104} = 1,26;$$

$$e_{105} = 1,75;$$

$$e_{106} = 1,02;$$

$$e_{107} = -1,73;$$

$$e_{108} = -4,15;$$

$$e_{109} = -12;$$

$$e_{110} = -10,6;$$

$$e_{111} = -9,76.$$

Горизонт прогнозирования  $n=10$ . Имея все необходимые данные рассчитаем MAE, MdAE, MSE, RMSE по формулам 2.12-2.15 соответственно:

$$MAE = 4,631$$

$$MdAE = 2,02$$

$$MSE = 38,57901$$

$$RMSE = 6,2112$$

После проведения расчетов можно сделать вывод, что данные виды оценок чувствительны к выбросам. Ошибка  $MSE$  сильно завышена, что

свидетельствует о ее ненадежности. Как можно заметить самый устойчивый результат имеет  $MdAE$ .

## 2. Процентные ошибки прогнозирования

Процентные ошибки рассчитываются на основании значений  $p_{102}-p_{111}$  (формула 2.16):

$$p_{102} = 0,13051;$$

$$p_{103} = 0,106421;$$

$$p_{104} = 0,109375;$$

$$p_{105} = 0,065469;$$

$$p_{106} = 0,114041;$$

$$p_{107} = 0,281165;$$

$$p_{108} = 0,836237;$$

$$p_{109} = 0,760948;$$

$$p_{110} = 0,760948;$$

$$p_{111} = 0,721893.$$

По формулам 2.17–2.20 рассчитаем ошибки  $MAPE$ ,  $MdAPE$ ,  $RMSPE$ ,  $RMdSPE(\%)$ :

$$MAPE = 32,02842$$

$$MdAPE = 12,22757$$

$$RMSPE = 44,06842$$

$$RMdSPE = 12,25526$$

Как мы видим значительное влияние на оценку оказывают выбросы, что приводит к неадекватной оценке модели прогнозирования.

## 3. Симметричные ошибки прогнозирования

Для расчета симметричных ошибок используют показатели  $s_{102}-s_{111}$ , рассчитанные по формуле 2.21:

$$s_{102} = 0,069811;$$

$$s_{103} = 0,056201$$

$$s_{104} = 0,039924;$$

$$s_{105} = 0,057851;$$

$$s_{106} = 0,033942;$$

$$s_{107} = 0,053944;$$

$$s_{108} = 0,123255;$$

$$s_{109} = 0,29484;$$

$$s_{110} = 0,275611;$$

$$s_{111} = 0,265217.$$

На основании полученных данных можем вычислить ошибки вида:  $sMAPE$ ,  $sMdAPE$ ,  $msMAPE$  (формулы 2.22–2.24):

$$sMAPE = 25,40995;$$

$$sMdAPE = 12,7662;$$

$$msMAPE = 0,235168.$$

Как и выше рассмотренные ошибки, симметричные являются крайне чувствительными к выбросам. Так же возникает проблема деления ноль, если прогнозное и реальное значение равны, но имеют противоположные знаки.

#### 4. Ошибки основанные на оценки прогноза базовых моделей

Базовой переменной для расчетов является критерий  $r_{102}-r_{111}$  (формула 2.25):

$$r_{102} = - 5,4878;$$

$$r_{103} = - 4,2618;$$

$$r_{104} = -3,07317;$$

$$r_{105} = - 4,26829;$$

$$r_{106} = - 2,42857;$$

$$r_{107} = - 4,21951;$$

$$r_{108} = - 10,122;$$

$$r_{109} = - 29,2683;$$

$$r_{110} = - 25,2381;$$

$$r_{111} = - 23,8049.$$

В данную группу ошибок входят следующие ошибки:  $MRAE$ ,  $MdRAE$  (формулы 2.27–2.28):

$$MRAE = 11,21725$$

$$MdRAE = 4,878049.$$

Основным недостатком оценок этого типа является деление на 0.

#### 5. Масштабированные ошибки прогнозирования

Базисным критерием оценок является величина  $q_{102}-q_{111}$  рассчитанными по формуле 2.29:

$$q_{102} = 5,443548;$$

$$q_{103} = 4,330645;$$

$$q_{104} = 3,048387;$$

$$q_{105} = 4,233871;$$

$$q_{106} = 2,467742;$$

$$q_{107} = 4,185484;$$

$$q_{108} = 10,04032;$$

$$q_{109} = 29,03226;$$

$$q_{110} = 25,64526;$$

$$q_{111} = 23,6129.$$

В данную группу входят следующие оценки:  $MASE$ ,  $MdASE$ ,  $RMSSE$  (формулы 2.30–2.32):

$$MASE = 11,20403$$

$$MdASE = 4,887097$$

$$RMSSE = 15,0271$$

Как и в оценках описанных выше, возникает ошибка деления на ноль.

#### 6. Относительные оценки

В данную группу входят оценки, рассчитываемые на основе отношения ошибок, рассмотренных для анализируемой модели прогнозирования и базовой модели. Так как в нашей работе базовая модель не рассмотрена, относительные оценки нами не рассмотрены.

#### 7. Другие группы оценок

В данной группе рассчитаны ошибки  $nRMSE$ ,  $inRSE$ ,  $Std\_AE$ ,  $Std\_APE$  (2.38–2.32):

$$nRMSE = 0,04039$$

$$inRSE= 5,236355$$

$$Std\_AE= 4,55955$$

$$Std\_APE= 33,424283$$

В качестве недостатка можно указать деление на нуль, в случае равенства нулю фактора нормализации.

### **3.4. Анализ приведенных в литературе рекомендаций к выбору оценок прогнозирования качества временных рядов**

Проблема прогнозирования, вследствие быстрых и трудно предсказуемых изменений внешней среды, является особенно сложной. С учетом этих трудностей и ошибок в прогнозах, возникает необходимость разработки рекомендаций к выбору оценки качества прогнозирования, сделанных на основании проанализированной литературы и расчета ошибок прогнозирования.

В связи с тем, что каждый критерий обладает недостатками, которые могут приводить к неточному определению качества прогнозирования, невозможно выделить один универсальный критерий.[4].

1. Поскольку проведенные расчеты ошибок оценок прогнозирования согласуются с предложенными в литературе и на стадии обработки данных не было выявлено аномалий, то наиболее устойчивыми оценками являются *MAE*, *MdAE*, *RMSE*;

2. Так как область изменения значений лежит в положительной полуплоскости, и данные не содержат выбросы, то целесообразно применять оценки вида: *sMAPE*, *sMdAPE*, *msMAPE*, что полностью совпадает с рекомендациями приведенными в литературе;

3. Данные нашего набора не содержат выбросы, и временной промежуток на котором производились расчеты не является настолько большим, чтобы не иметь близких по значению данных, в связи с этим целесообразность использования масштабированных оценок необходимо проверить используя другую модель.

4. Поскольку модель численности нетрудоустроенных граждан имеет ярко выраженные сезонные колебания, что целесообразно использовать нормализованные оценки ошибок прогнозирования. При этом фактор нормализации должен быть равен величине сезона;

5. В случае если не имеется предварительного анализа, о качестве сведений необходимо анализировать уже полученные результаты. Также требуется включение различных оценок ошибок прогнозирования. После расчета требуется проанализировать: возникновение ситуации деление на 0, а так же возникновение противоречий. Вы деленные противоречия согласуются с литературой:

- Результаты прогнозирования двух моделей, не совпадают по разным оценкам ошибок;
- На разных наборах данных получаются разные результаты оценок ошибок прогнозирования.

## ЗАКЛЮЧЕНИЕ

Для оценки качества прогнозирования используются различные критерии: оценки ошибок прогнозирования, скорость расчета, интерпретируемость и другие. Оценка ошибки прогнозирования или точности прогноза являются наиболее важными при решении практических задач.

В работе проведен анализ данных временного ряда численности нетрудоустроенных граждан, состоящих на учете в государственных учреждениях службы занятости с января 2008 по март 2017 года

Анализируя временный можно утверждать, что:

- имеется тенденция к уменьшению числа нетрудоустроенных граждан (линейный тренд с отрицательным углом наклона), обусловленная положительными изменениями социальной ситуации.
- имеются сезонные колебания с максимумом в зимние периоды и минимумом в летние периоды.

Так же в работе осуществлен расчет основных видов оценок прогнозирования:

- Абсолютные ошибки прогнозирования;
- Процентные ошибки прогнозирования;
- Симметрические ошибки прогнозирования;
- Ошибки, основанные на оценке прогноза базовых моделей;
- Масштабированные ошибки прогнозирования;
- Относительные оценки;
- Другие группы оценок.

На основании этих расчетов и проанализированной литературы получены рекомендации к выбору оценки качества прогнозирования по модели численности нетрудоустроенных граждан Алтайского края состоящих на учете в государственных учреждениях службы занятости.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Armstong J.S., Collopy F. Error measures for generalizing about forecasting methods: Empirical comparisons. – 1992. - №8 (1). – P. 69-80.
2. Hyndman R.J., Koehler A.B. Another look at measures of forecast accuracy // *International Journal of Forecasting*. – 2006. – № 22(4). – P. 679-688.
3. Theodosiou M. Forecasting monthly and quarterly time series using STL decomposition – 2011. – №27(4). – P. 1178-1195.
4. Mahmoud E. *Accuracy in forecasting: A survey* // *Journal of Forecasting*. – 1984. – № 3(2). – P. 139-159.
5. Амосов, О.С. Интеллектуальные информационные системы. Нейронные сети и нечеткие системы: Учебное пособие. – Комсомольск-на-Амуре: ГОУВПО «КНАГТУ», 2004. – 104 с.
6. Бассовский Л. Е. Прогнозирование и планирование в условиях рынка: Учебное пособие. – М.: ИНФРА, 2009. – 259 с.
7. Басовский Л. Е. Прогнозирование и планирование в условиях рынка. Учеб. пособие.- М. : ИНФРА-М, 2011
8. Белов А.А., Проскуряков А.Ю. Сглаживание временных рядов на основе вейвлет- преобразования в системах автоматизированного экологического мониторинга // *Методы и устройства передачи и обработки информации*. 2010. №12. С 21-24.
9. Белов А.А., Кропотов Ю.А., Проскуряков А.Ю. Автоматизированный анализ и обработка временных рядов данных о загрязняющих выбросах в системе экологического контроля // *Информационные системы и технологии*, 2010. – №6(62). – С. 28– 35.
10. Белов А.А., Проскуряков А.Ю. Вопросы обработки экспериментальных временных рядов в электронной системе автоматизированного контроля
11. Владимирова Л. П., Прогнозирование и планирование в условиях рынка: Учебное пособие. 2-е изд., перераб. и доп. – М.: Дашков и К, 2008. – 307 с.



12. Зайдель А.Н. Элементарные оценки ошибок измерений. – 2-е изд. испр. и доп. – Л.: «Наука», Ленинградское отд., 1967. – 88с.
13. Кацко, И.А. Практикум по анализу данных на компьютере. Под ред. Г.В.Гореловой. – М.: КолосС, 2009. – 278с.
14. Каллан Р., Основные концепции нейронных сетей: Учебное пособие – М.: Вильямс, 2001. – 288 с.
15. Каллан Р., Основные концепции нейронных сетей: Учебное пособие – М.: Вильямс, 2001. – 288 с.
16. Козьева А., Вертакова Ю. В., Бабич Т. Н. Прогнозирование и планирование в условиях рынка. Учеб. пособие для вузов. – М. : ИНФРА-М, 2012.
17. Комашинский В.И., Смирнов Д.А. Нейронные сети и их применение в системах управления и связи: Учебное пособие. – М.: 2002. – 94 с.
18. Короткий С.А., Нейронные сети Хопфилда: Учебное пособие. – М.: ЮНИТИ-ДАНА, 2005. – 315 с. 8. Роберт Каллан. Основные концепции нейронных сетей = The Essence of Neural Networks First Edition. – 1-е. – «Вильямс», 2001. – 288с.
19. Меламед М.А. Современные методы анализа и прогнози-рования режимов электропотребления в электроэнергетических системах // Итоги науки и техники. – Сер. Энергетические системы и их автоматизация. – 1988. – Т.4. – С. 4–11.
20. Тихонов Э.Е. Прогнозирование в условиях рынка. Невинномысск, 2006. 221 с.
21. Томпсон А.А., Дж. Стрикленд Ш. А. Стратегический менеджмент: Концепции и ситуации: Учебное пособие для вузов. – М.: ИНФРА-М, 2010. – 411 с.
22. Тренев Н. Н., Стратегическое управление: Учебное пособие. М

23. Саркисян С. А. Теория прогнозирования и принятия решений / С.А. Саркисян, В.И. Каспин, В.А. Лисичкин, Э.С. Минаев, Г.С. Пасечник. – М.: Высшая школа, 1977. – 351 с.
24. Щербаков М. В., Щербакова Н. Л., Козлов И. П., Камаев В. А. Коннективистские модели идентификации динамики систем на коротких интервалах наблюдения с заданным множеством классов поведения // Известия ЮФУ. Технические науки. – 2012. – №1. – С. 83-91.
25. Е.М. Четыркин. Статистические методы прогнозирования. – М.: Статистика, 1977. – 199 с.
26. В. Черняк. Модель экономики: выбор Украины. // Экономика Украины, 1995, №9.
27. А.А.Френкель. Прогнозирование производительности труда: методы и модели. – М.: Экономика, 1989. – 214 с.