

ISSN 2500-316X (Online)

<https://doi.org/10.32362/2500-316X-2020-8-4-7-45>



УДК 004.942, 004.048, 001.51, 519.688

**Обзор современных моделей и методов анализа  
временных рядов динамики процессов  
в социальных, экономических  
и социотехнических системах**

**Е.Г. Андрианова<sup>1,®</sup>,  
С.А. Головин<sup>1</sup>,  
С.В. Зыков<sup>2</sup>,  
С.А. Лесько<sup>1</sup>,  
Е.Р. Чукалина<sup>1</sup>**

<sup>1</sup>МИРЭА – Российский технологический университет, Москва 119454, Россия

<sup>2</sup>НИУ «Высшая школа экономики», Москва 101000, Россия

<sup>®</sup>Автор для переписки, e-mail: andrianova@mirea.ru

Описаны направления перспективных исследований в области анализа и моделирования динамики временных рядов процессов в сложных системах с присутствием человеческого фактора. Динамика процессов в таких системах описывается нестационарными временными рядами. Прогнозирование эволюции подобных систем имеет большое значение для управления процессами в социальных (избирательные кампании), экономических (фондовые, фьючерсные и сырьевые рынки) и социотехнических системах (социальные сети). Приведены общие сведения о временных рядах и задачах их анализа. Рассмотрены современные методы анализа временных рядов для экономических процессов. Результаты показывают, что экономические процессы нельзя считать полностью случайными, так как они имеют тенденцию к самоорганизации и, кроме того, подвержены влиянию памяти о предыдущих состояниях. Выявлено, что одной из главных задач при моделировании процессов в социотехнических системах (например, социальных сетях) является разработка математического аппарата для приведения данных к единой

шкале измерений. Проанализированы современные модели анализа и прогнозирования электоральных процессов на основе анализа временных рядов: структурные, опросные, гибридные. На основе проведенного анализа рассмотрены их достоинства и недостатки. В заключении подтвержден вывод, что для описания процессов в сложных системах с наличием человеческого фактора помимо традиционных факторов, необходимо разрабатывать и использовать методы и инструментальные средства для учета возможности самоорганизации человеческих групп и наличия памяти о предыдущих состояниях системы.

**Ключевые слова:** нестационарные временные ряды, сложные системы, анализ экономических и социальных процессов, анализ социальных сетей, прогнозирование эволюции временных рядов, модели и методы анализа временных рядов, человеческий фактор, электоральные процессы, самоорганизация системы.

*Для цитирования:* Андрианова Е.Г., Головин С.А., Зыков С.В., Лесько С.А., Чукалина Е.Р. Обзор современных моделей и методов анализа временных рядов динамики процессов в социальных, экономических и социотехнических системах. *Российский технологический журнал*. 2020;8(4):7-45. <https://doi.org/10.32362/2500-316X-2020-8-4-7-45>

## **Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems**

**Elena G. Andrianova<sup>1,@</sup>,  
Sergey A. Golovin<sup>1</sup>,  
Sergey V. Zykov<sup>2</sup>,  
Sergey A. Lesko<sup>1</sup>,  
Ekaterina R. Chukalina<sup>1</sup>**

<sup>1</sup>MIREA – Russian Technological University, Moscow 119454, Russia

<sup>2</sup>NRU "Higher School of Economics", Moscow 101000, Russia

@Corresponding author, e-mail: andrianova@mirea.ru

The directions of perspective research in the field of analysis and modeling of the dynamics of time series of processes in complex systems with the presence of the human factor are described. The dynamics of processes in such systems is described by non-stationary time series. Predicting the evolution of such systems is of great importance for managing processes in social (election campaigns), economic (stock, futures and commodity markets) and socio-technical systems (social networks). The general information on time series and tasks of their analysis is given. Modern methods of time series analysis for economic processes are considered. The results show that economic processes cannot be considered completely random, since they tend to self-organize and, moreover, are subject to the influence of memory of previous states. It was revealed that one of the main tasks in modeling processes in sociotechnical systems (for example, social networks) is the development of a mathematical apparatus for bringing data to a single measurement scale. The modern models of analysis and forecasting of electoral processes based on the analysis of time series: structural, polling, hybrid. Based on the analysis, their advantages and disadvantages are considered. In conclusion, it was concluded that to describe processes in

complex systems with the presence of the human factor, in addition to traditional factors, it is necessary to develop and use methods and tools to take into account the possibility of self-organization of human groups and the presence of memory about previous states of the system.

**Keywords:** non-stationary time series, complex systems, analysis of economic and social processes, analysis of social networks, forecasting the evolution of time series, models and methods of analysis of time series, human factor, electoral processes, self-organization of the system.

**For citation:** Andrianova E.G., Golovin S.A., Zykov S.V., Lesko S.A., Chukalina E.R. Review of modern models and methods of analysis of time series of dynamics of processes in social, economic and socio-technical systems. *Rossiiskii tekhnologicheskii zhurnal = Russian Technological Journal*. 2020;8(4):7-45 (in Russ.). <https://doi.org/10.32362/2500-316X-2020-8-4-7-45>

## Введение

Анализ динамики временных рядов и прогнозирование их эволюции имеют большое значение для управления различными процессами в социальных (например, избирательные кампании), экономических (фондовые, фьючерсные и сырьевые рынки) и социотехнических системах (например, социальные сети).

Следует отметить, что сложность наблюдаемых процессов существенным образом возрастает при переходе от технических к социотехническим и далее – к социальным системам. Многие технические системы по непредсказуемости наблюдаемых процессов уступают системам, в которых одним из элементов является человек. Если для прогнозирования динамики процессов в технических системах основной проблемой является наличие существенной стохастичности, то в социальных, экономических и социотехнических системах появляется человеческий фактор.

С одной стороны, появление человеческого фактора приводит к стохастичности свойств и неопределенности в социальных, экономических и социотехнических системах. С другой стороны, наличие человеческого фактора создает предпосылки для самоорганизации системы и определяет существование памяти о её предыдущих состояниях. Всё это приводит к появлению организованной сложности (эмерджентности). Возникновение эмерджентности является результатом появления системных связей и адаптивного перераспределения функций между элементами системы. Таким образом динамика процессов в социальных, экономических и социотехнических системах описывается нестационарными временными рядами, а попытки представить их в виде квазистационарных участков могут приводить к неправильному описанию наблюдаемых процессов и неверному построению управляющих прогнозов.

При описании эволюции временных рядов важным моментом является определение так называемых точек «разладки» (в экономике это называется «точкой смены тренда»). Поясним этот эффект на примере технических систем. В процессе своего функционирования технические системы могут испытывать случайные широкополосные и сложно-периодические возмущения (так называемые ритм-сигналы), не подпадающие ни под один из известных законов распределения случайных величин, что в некоторых случа-

ях может привести к аномальному состоянию технической системы («разладке»). Большинство традиционных методов и моделей для выявления эффекта «разладки» основываются на предположении, что вероятностные характеристики выходного ритм-сигнала у исправного технического объекта с течением времени не изменяются, т.е. измеренный стохастический сигнал является стационарным. В таких методах обычно используют преобразование Фурье (Fast Fourier Transform – FFT), распределение Вигнера – Вилля, вейвлет-преобразование или преобразование Гильберта – Хуанга. Если ритм-сигналы носят сложный нестационарный характер, тогда применение традиционных методов выявления разладок в технических системах невозможно и необходимы новые методы анализа динамики нестационарных временных рядов. Следует отметить, что технические системы при всей сложности реализующихся в них процессов являются более простыми и детерминированными по отношению к экономическим, и тем более социальным или социотехническим системам, поскольку в них все основные характеристики процессов могут быть измерены числовыми физическими величинами.

Одним из важнейших примеров и объектов социотехнических систем являются социальные сети. Важность изучения нестационарных временных рядов, описывающих процессы в социотехнических системах, заключается в том, что в последние несколько лет наблюдается значительный рост влияния событий, происходящих в интернете, как на политические, так и на социальные процессы в мире. Одним из примеров успешного использования социальных сетей можно назвать предвыборную кампанию Барака Обамы в США в 2007 и 2008 годах [1]. Благодаря революционным методам мобилизации социальных сетей была получена поддержка огромной аудитории и собраны необходимые деньги. Ещё одним примером может послужить предвыборная кампания в Израиле в 2015 году [2], когда кандидаты, стремясь привлечь избирателей, активно пользовались социальными сетями, проводя различные акции. В одном из видеороликов Биньямин Нетаньяху, нынешний премьер-министр, выступил в роли детской сиделки, намекая, что никто лучше него не обеспечит безопасность детей в Израиле.

Актуальность проведения теоретических и практических исследований процессов, описываемых нестационарными временными рядами в социальных, экономических и социотехнических системах, обуславливается необходимостью разработки методик прогнозирования поведения людей в обществе, возникновения новых запросов и потребностей, а также предотвращения негативных настроений. Особенно важным это становится в связи с появлением и развитием новых информационных технологий. В частности, события последнего времени показывают невозможность введения полного контроля над интернетом и социальными сетями. В частности, компания SpaceX анонсировала технологию дешевого спутникового доступа в интернет практически с любого мобильного устройства. Эта технология обеспечивает неконтролируемую национальными правительствами передачу данных [3]. Кроме того, необходимо учитывать и возможности использования так называемых анонимайзеров (обеспечивают анонимность выхода в интернет) и методов шифрования данных для затруднения обнаружения действий пользователей сетей.

## 1. Свойства и модели временных рядов

Временным рядом  $\{x_i\}_{i=1}^n$  называется массив из  $n$  чисел, представляющих собой значения некоторой наблюдаемой динамической переменной  $x(t)$  с некоторым постоянным шагом  $\tau$  по времени,  $t_i = t_0 + (i-1)\tau$ ;  $x_i = x(t_i)$ ,  $i = 1, \dots, n$ .

Таким образом, любой временной ряд имеет два обязательных элемента: время и уровень ряда (конкретное значение показателя). Их можно различать по следующим признакам:

- время (моментное, интервальное);
- форма представления уровней (временные ряды абсолютных, относительных и средних величин);
- расстояние между интервалами времени (полные и неполные временные ряды);
- содержание показателей (временные ряды частных и агрегированных показателей).

Для корректного формирования временных рядов наиболее важным параметром является сопоставимость уровней, образующих их.

В теории временных рядов выделяют две основные задачи: идентификации и прогноза.

Методы прогнозирования разделяют на две группы [4]: интуитивные и формализованные. Первая группа методов не предполагает разработку моделей прогнозирования и отражает суждения экспертов. Данная группа методов применяется, когда объект слишком прост или настолько сложен, что невозможно аналитически учесть влияние внешних факторов. Вторая группа методов основывается на построении модели прогнозирования.

Наиболее активно используемыми при моделировании сложных процессов являются регрессионные модели. Регрессионные модели прогнозирования представляют собой функцию независимой переменной и параметров с добавленной случайной переменной. Простая линейная регрессия описывается следующим уравнением:

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

где  $\alpha_0$ ,  $\alpha_1$  – коэффициенты регрессии,  $\varepsilon_t$  – ошибка модели.

Целью анализа является выявление зависимости между исходной переменной и внешними факторами. Для определения коэффициентов регрессии используются метод наименьших квадратов и метод максимального правдоподобия. Помимо простой линейной регрессии используют множественную и нелинейную регрессию.

Авторегрессионные модели прогнозирования основываются на модели временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда.

Наиболее популярными являются модель авторегрессии

$$Z(t) = C + \phi_1 Z(t-1) + \phi_2 Z(t-2) + \dots + \phi_n Z(t-p) + \varepsilon_t,$$

где  $C$  – вещественная константа,  $\phi_1, \phi_2, \dots, \phi_n$  – коэффициенты,  $p$  – порядок авторегрессии,  $\varepsilon_t$  – ошибка модели,

и модель скользящего среднего:

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

где  $q$  – порядок скользящего среднего,  $\varepsilon_t$  – ошибка прогнозирования.

Часто в одной модели объединяют модели авторегрессии и скользящего среднего. Такая объединенная модель представляет собой фильтр в виде скользящего среднего порядка  $q$  и авторегрессию фильтрованных значений процесса порядка  $p$ .

В моделях экспоненциального сглаживания (взвешенное скользящее среднее, экспоненциальное сглаживание, модель Хольта, модель Хольта – Винтерса) выявление и анализ тенденции временного ряда часто производится с помощью его выравнивания или сглаживания. В основу экспоненциального сглаживания заложена идея постоянного пересмотра прогнозных значений по мере поступления фактических.

Функция модели имеет вид:

$$Z(t) = S(t) + \varepsilon_t,$$

$$Z(t) = \alpha \cdot Z(t-1) + (1 - \alpha) \cdot S(t-1),$$

где  $\alpha$  – коэффициент сглаживания, меняющийся от нуля до единицы, а  $S(1) = Z(0)$  – начальные условия. В данной модели каждое последующее сглаженное значение  $S(t)$  является взвешенным средним между предыдущим значением временного ряда  $Z(t)$  и предыдущим сглаженным значением  $S(t-1)$ . Такие модели наиболее подходят для долгосрочного прогнозирования.

При помощи модели на нейронных сетях возможно моделирование нелинейной зависимости будущего значения временного ряда от его фактических значений и от значений внешних факторов. Модель на основе искусственных нейронных сетей, состоящих из нейронов, описывается уравнениями:

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

$$Z(t) = \phi(U(t)),$$

где  $Z(T-1), \dots, Z(t-m)$  – входные сигналы,  $\omega_1, \dots, \omega_m$  – синаптические веса нейрона,  $b$  – порог,  $\phi(U(t))$  – функция активации (функция единичного скачка, кусочно-линейная функция, сигмоидальная функция).

Существуют также модели на цепях Маркова, на классификационно-регрессионных деревьях, на основе генетического алгоритма, на опорных векторах, на основе передаточных функций, на нечеткой логике и т.д. Также отметим, что на сегодняшний день существует множество модификаций названных моделей.

Из-за нелинейности и изменчивости характеристик нестационарных временных рядов процессов, протекающих в сложных социальных, экономических и социотехнических системах, традиционные методы анализа и моделирования, например, такие как интегрированная модель авторегрессии – скользящего среднего (ARIMA,

модель Бокса – Дженкинса [4]) и многие другие часто приводят к неточным или ошибочным результатам.

Традиционный подход к анализу нестационарных временных рядов основан на том, чтобы с помощью применения линейных методов свести их к стационарным временным рядам, например упомянутые модели ARIMA.

Эти модели оперируют не с функциями распределения, а непосредственно с элементами временного ряда. Ряды, не укладывающиеся в рамки регрессионного анализа, чаще всего изучаются различными адаптивными эвристическими методами, не имеющими четкого математического обоснования. В них предполагается, что ряды на некоторой длине могут быть описаны стационарной моделью типа регрессии или авторегрессии, а параметры модели могут быть пересчитаны с учетом новой информации или с учетом сравнения предсказанного значения с фактом. Недостатком этих подходов является то, что длина участка возможной стационарности является неизвестной величиной и в любой момент времени может произойти смена тренда (разладка).

Задача прогнозирования в стационарном случае может быть сведена к задаче аппроксимации средних величин. Для стационарных случайных процессов выборочное распределение сходится по вероятности к теоретическому. Существуют независимые от времени конечные моменты теоретического распределения нескольких первых порядков (математическое ожидание, дисперсия, асимметрия, эксцесс) и отклонения выборочных моментов от их теоретических значений распределены асимптотически нормально.

В настоящее время существует более тысячи статистических тестов (критериев), которые применяются для того, чтобы отнести изучаемый случайный процесс к тому или иному классу и использовать для его описания определенную математическую модель. В прикладных задачах часто используется критерий согласия Колмогорова (1933 г.) для определения близости выборочной функции распределения случайной величины к стационарному распределению. В дальнейшем на его основе были получены другие широко применяемые асимптотические критерии [5–9] о принадлежности двух выборочных распределений одной генеральной совокупности: критерий Колмогорова – Смирнова (1939 г.), Вальда – Волфовица (1940 г.), Вилкоксона (1945 г.), Манна – Уитни (1947 г.), Гнеденко – Королюка (1951 г.), критерий тренда в среднем значении Аббе – Линника [10, 11], критерий тренда в дисперсии Фостера – Стюарта [10, 12] и другие критерии, применяемые к оценкам выборочных моментов (Стьюдента, Фишера, Крамера – Уэлча, «омега-квадрат» и др.) [13].

Доверительные вероятности и доверительные интервалы для выборочных оценок параметров стационарных распределений могут быть получены на основе теоремы Вольда (1938 г.), согласно которой любой стационарный случайный процесс представляется в виде суперпозиции детерминированного процесса и белого шума, и теоремы Гофдинга (1948 г.), которая позволяет определить скорость сходимости по вероятности выборочных моментов и вероятность отклонения их значений от теоретических (если моменты генеральной совокупности для стационарной случайной величины распределены асимптотически нормально).

Перечисленные утверждения позволяют представить стационарный временной ряд в виде суммы детерминированной составляющей (трендовой или периодической)

и остатка, автокорреляционная функция которого с достаточной точностью близка к нулю, что свидетельствует о близости остатка к белому шуму. После этого ставится задача о нахождении наиболее близкой статистики (функции распределения), моделирующей поведение остатка. В адаптивных методах описания априори не известно, является ли ряд стационарным или нет, а также по выборке какого объема следует находить, например, скользящее среднее, чтобы получить наименьшую ошибку прогноза.

Для нестационарных временных рядов индикаторы тех или иных его свойств имеют свой специфический вид, не обобщаемый на ряды другого типа. Например, индикатор линейного тренда не особенно эффективен для рядов с квазипериодическим изменением, как и индикатор нестационарности дисперсии для рядов с квазилинейным трендом [14]. Более того, индикаторы, основанные на некоторых средних характеристиках ряда (например, несколько первых моментов), не образуют базисной системы, по которой можно определить тенденцию локального по времени изменения случайного процесса.

Корректно оценить вероятность того или иного значения функционала от генеральной совокупности по одной выборке (интервалу реализации временного ряда) можно только для стационарных процессов. Однако во многих практических задачах при большом числе наблюдений за случайным процессом, осуществляемым в скользящем окне, обнаруживается, что если процесс не является стационарным, то число ошибок в принятии той или иной статистической гипотезы оказывается в разы больше, чем уровень значимости, на котором по классическому критерию принималось решение.

В случае нестационарных временных рядов все асимптотические критерии, гарантирующие увеличение точности аппроксимации с увеличением объема выборки, оказываются несостоятельными. Аналогичные проблемы возникают и при использовании сглаженного скользящего усреднения. Если ряд нестационарный, то средние (скользящие, «растущие» – т.е. взятые по выборке растущего объема, или любые другие) не являются состоятельными оценками моментов распределения, так как сходимости по вероятности в общем случае нет.

В нестационарном случае отсутствует само понятие генеральной совокупности, что делает неприменимым аппарат математической статистики, кроме тех случаев, когда априори известна функциональная принадлежность модели процесса к классу распределения, что на практике, как правило, неизвестно.

При исследовании стационарных случайных процессов, согласно теореме Гливенко (о сходимости эмпирической вероятности к теоретическому распределению) [15], чем больше будет учтено наблюдаемых значений, тем точнее могут быть получены теоретические характеристики распределения случайной величины из определенного промежутка. Для нестационарных случайных процессов данное условие, в силу их специфики, не может быть выполнено, что затрудняет возможность использования результатов их анализа для дальнейшего прогнозирования. Нестационарность не позволяет получить достаточно большое число значений величины временного ряда для построения генеральной совокупности. При анализе стационарного процесса полученные значения математического ожидания и дисперсии случайной величины могут быть использованы для построения прогноза на любых интервалах времени наблюдения, в случае нестационарных временных рядов возникает момент разладки, после которого полученные ранее ха-

характеристики распределения не могут давать точный прогноз дальнейшего поведения временного ряда.

В работах [7, 13, 16, 17] приведены непараметрические критерии оценивания данных, использующие методы математического моделирования Монте-Карло [17, 18]. Однако, следует отметить, что эти методы применимы только к стационарным распределениям и не могут корректно использоваться для анализа нестационарных временных рядов.

Например, в работе [18] для моделирования траекторий некоторого случайного процесса со скачками, применительно к ценам акций предприятий авиационной отрасли РФ, используются стохастические дифференциальные уравнения с зависящими от времени коэффициентами сноса и диффузии и эрланговский поток событий для описания скачков, но практическая реализация случайных траекторий основана на стационарном методе Монте-Карло при постоянных коэффициентах стохастического уравнения.

Учитывая всё вышесказанное, можно сделать вывод, что необходим поиск новых методов анализа динамики сложных систем или новых подходов к описанию нестационарных временных рядов, особенно если возможна самоорганизация таких систем и наличие памяти о предыдущих состояниях, например, моделей и методов, основанных на различных видах или сочетаниях алгоритмов машинного обучения, таких как нейронные сети [19–22], нечеткая логика [23], метод опорных векторов регрессии [24], наборы правил на основе генетического программирования сетевых приложений [25] и ряда других.

## **2. Моделирование экономических процессов на основе анализа временных рядов**

Следует отметить, что наиболее хорошо область анализа временных рядов была проработана для экономических процессов (в силу заинтересованности большим числом их участников в получении положительных результатов).

В статье [19] описан новый подход к фундаментальному анализу данных нестационарных временных рядов для выявления связей между поведением рынка и внешней информацией, выдвинута гипотеза о возможной логической актуальности некоторой внешнеэкономической информации и ценовых движений валютной пары USD/GBP. Авторы [19] с использованием нейронных сетей провели ряд экспертиз с целью определения существования правдоподобных связей между внешними источниками информации и ценовыми движениями валютной пары. Авторы [19] применяют два класса моделей искусственного интеллекта: нейронную сеть Psi Sigma (PSI) и эволюционный алгоритм Gene Expression (GEP). Было показано, что PSI способна фиксировать корреляции более высокого порядка в наборе данных, а GEP определяет различия между генотипом и фенотипом. Это позволяет получить более точное прогнозирование по сравнению с «традиционными» моделями. В работе [20] утверждается, что нейронные сети хорошо подходят для моделирования движения цен и дают возможность извлекать информацию из больших объемов данных, которая необходима для анализа сложных сигналов о движении финансовых индексов и цен.

Авторы [20] применяют как параметрические (нейронные сети с активными нейронами), так и непараметрические (аналоговые комплексобразующие) методы самоорганизующегося моделирования для ежедневного прогнозирования валютного рынка. В работе предлагается комбинированный подход, при котором параметри-

ческие и непараметрические методы самоорганизации объединяются последовательно, используя преимущества отдельных методов с целью повышения их производительности. Сначала рассматривается параметрический подход для нейронных сетей с активными нейронами. Затем обсуждается непараметрический метод самоорганизующегося моделирования, известный как аналоговое комплексирование (Analog Complexing) и, наконец, описывается новый гибридный метод, вытекающий из последовательной комбинации двух предыдущих. Авторы показали, что комбинированный метод дает хорошие результаты и превосходит отдельные при тестировании с двумя обменными курсами: американский доллар и немецкая марка по отношению к британскому фунту.

В работе [21] рассматривается методология проектирования и тестирования биржевых торговых систем (роботов), использующих технологии мягких вычислений (нечеткой логики) и искусственные нейронные сети (ANN). Представленная в работе методология отделяет процесс обучения нейронных сетей и выбора параметров процесса невыборочного бенчмаркинга. Для обучения и тестирования ANN данные должны быть логически (или физически) разделены как минимум на два набора – набор обучения и набор тестирования. Основной принцип заключается в том, чтобы охватить как можно более разнообразную рыночную активность (с длительным обучением), сохраняя при этом как можно более длинное окно тестирования (чтобы увеличить уверенность в модели).

В работе [24] на основе теории хаоса представлена непараметрическая модель для прогнозирования будущего поведения валютных курсов. Теория хаоса предполагает, что поведение финансовых рынков хаотично и обеспечивает основу для учета динамики нелинейных систем, которые могут выявлять скрытые шаблоны и тенденции в финансовых данных, и которые, в свою очередь, не могут быть получены обычными статистическими методами. Предложенная двухступенчатая модель сочетает в себе фазовое пространство, реконструированное из одномерных хаотических временных рядов в комбинации с векторной регрессией. На первом этапе метод вложения координат задержки преобразует необработанные временные ряды в фазовое пространство, подходящее для прогнозирования. На втором этапе стандартная векторная регрессия поддержки (SVR) применяется к преобразованной серии для окончательного прогнозирования. Радиальная базисная функция используется как ядро SVR. Полученная модель, названная «Chaos – SVR», превосходит обычную SVR, нейронную сеть обратного распространения («BPNN») и ее комбинацию с преобразованным фазовым пространством («Chaos – BPNN») на общих метриках для измерения производительности модели, таких как MSE (средняя квадратичная ошибка), RMSE (среднее значение квадрата ошибки) и MAE (средняя абсолютная ошибка).

Однако следует отметить, что в реальности биржевые рынки обладают свойством самоорганизации и памятью о предыдущих событиях, т.е. не могут быть описаны в полной мере теорией хаоса.

В работе [26] представлен новый подход к анализу нечетких временных рядов высокого порядка. Авторы использовали адаптивный упорядоченный нечеткий временной ряд (FTS) для прогнозирования на валютном рынке FOREX. Исходный ряд с помощью быстрой самоорганизующейся стратегии разбивается на неравные интервалы, а выбор порядка ряда осуществляется с помощью адаптивного алгоритма. Предложенная модель

анализа состоит из двух этапов. На первом – осуществляется использование самоорганизующихся карт (SOM), которые позволяют осуществить более быструю кластеризацию данных по сравнению с генетическими алгоритмами (GA). На втором – производится адаптивный выбор порядка ряда, для чего используется метод трех агентов, которые являются агентом голосования, статистическим агентом и эмоциональным агентом. Эти агенты работают последовательно, чтобы найти наилучший порядок высокоуровневых нечетких временных рядов. Агент по голосованию пытается достичь консенсуса между различными решениями. При выборе порядка нечеткого временного ряда используются статистический анализ и методы принятия эмоциональных решений. Сравнение предложенного метода с более ранними исследованиями демонстрирует улучшенную точность прогнозирования при одинаковых затратах на вычисления.

В [27] авторы используют генетическое сетевое программирование (GNP) на основе правил для создания системы принятия решений на фондовом рынке. GNP с накоплением правил – это алгоритм, который извлекает большое количество правил из действий пользователей на протяжении поколений и хранит их в пулах правил, что отличается от общих эволюционных вычислений. GNP с накоплением правил можно рассматривать в качестве эволюционирующих генераторов правил на этапе обучения, а созданные правила в пулах правил – использовать для принятия решений. Большое количество правил купли-продажи извлекается из действий отдельных пользователей, а затем используется механизм классификации, чтобы правильно определить, покупать или продавать акции на основе извлеченных правил. Преимущество GNP заключается в том, чтобы запоминать различные ситуации, имевшие место в период обучения и хранить их в качестве правил в пулах правил. Однако технические аналитики считают, что в графике рынка есть закономерности, которые не могут быть обнаружены.

Следует особо подчеркнуть, что многие из описанных выше подходов являются очень хорошим инструментом, но качество их применения ограничивается несовершенством математических моделей, используемых для исследования динамики формирования временных рядов.

Анализ и прогнозирование поведения нестационарных временных рядов могут быть основаны на применении выборочных функций распределения. Временной ряд разделяется на отдельные области, в каждой из которых он является квазистационарным, со своей выборочной функцией распределения, а между областями имеется часть ряда, в которой происходит переходной процесс (разладка). Продолжительность переходного процесса определяется как факторами, характеризующими смену режима (собственно разладка), так и объемом выборки, который используется для проведения статистического анализа [13]. Параметры выборочной функции распределения могут быть установлены на основе анализа данных, наблюдаемых на заданном интервале времени. На практике необходимо решить две задачи. Во-первых, определить интервал времени квазистационарности. Во-вторых, понять наступление разладки в течение переходного периода, причем с минимальным запаздыванием. Следовательно, объем выборки должен быть заведомо меньше длительности переходного периода, но в то же время достаточным для того, чтобы оценки вероятностного распределения были достоверны [13]. Таким образом, возникает самосогласованная оптимизационная задача по определению индикатора разладки с наибольшей достоверностью.

Недостатком многих подходов является то, что они применяются к единственной реализации случайного процесса, тогда как для эволюционирующих распределений методически более корректно изучать ансамбль возможных траекторий. Если функция распределения нестационарна, то обучение алгоритма распознавания на прошлых данных часто оказывается несостоятельным. Имеется только одна траектория, которая в силу нестационарности не позволяет использовать большой объем выборки для тестирования тех или иных индикаторов локального поведения временного ряда с целью оценки вероятности их правильного срабатывания при выборе биржевой стратегии.

Для анализа нестационарных процессов можно использовать методы, основанные на анализе эволюции семейства выборочных распределений процесса, отличающихся объемами выборок. В работе [28] описана так называемая горизонтная статистика и изучены некоторые ее свойства для индикации смены уровня хаоса нестационарных временных рядов. Данный подход был далее развит в работе [13], где используются нестационарные статистики и определяются расстояния между выборочными плотностями функций распределения в различных нормах, получившие название согласованный уровень нестационарности, а также вводится понятие статистической добротности ряда для их согласования. По мнению авторов, эта статистика позволяет более точно определить моменты разладки.

Идентификация состояния нестационарного случайного процесса может быть сформулирована как задача распознавания выборочной функции распределения (ВФР) как принадлежащей определенной генеральной совокупности. Это может быть решено либо путем оценивания значений параметров распределения известного функционального вида, либо в рамках непараметрического подхода с использованием критерия Колмогорова – Смирнова [29].

На практике часто возникает задача тестирования функционала управления, заданного на траектории случайного процесса. Примером такого функционала является фондовый рынок, правила входа-выхода на который и их численные значения должны быть оптимизированы не по одной траектории, а по ансамблю траекторий. В результате необходимо также иметь генератор случайных величин, распределение которых эволюционирует, причем закон эволюции является непараметрическим.

Рассмотрим подход, при котором используется метод распознавания текущей ситуации как «благоприятной» для входа в рынок на основе усреднения выборочных траекторий нестационарного случайного процесса.

Индикаторы представляют собой определенные функционалы, заданные на фрагментах траектории случайного процесса. Для оценки эмпирической условной вероятности того, что определенный интервал значений индикатора отвечает ожидаемому поведению ряда, требуется сгенерировать набор возможных траекторий временного ряда, выборочная функция распределения которого эволюционирует ожидаемым образом. Затем проверить на сгенерированном наборе траекторий устойчивость срабатывания индикатора.

В качестве таких функционалов (индикаторов), например, можно использовать опционы, а достижение заданной величины доходности – как биржевой индикатор (критерий, при достижении которого осуществляется продажа или покупка актива).

Опционные контракты *call*-опционов  $C(t)$  и *put*-опционов  $P(t)$  связаны с функцией распределения  $p(x,t)$  следующим образом:

$$C(t) = \int_{x_s}^{+\infty} (x - x_s) \cdot p(x,t) dx \text{ и } P(t) = \int_0^{x_s} (x_s - x) \cdot p(x,t) dx,$$

где  $x_s$  – цена исполнения опциона (*strike price*),  $t$  – дата исполнения (*expiry date*),  $x$  – текущая стоимость,  $p(x,t)$  – плотность вероятности достижения стоимости  $x$  в момент времени  $t$  в качестве которой может быть использована восстановленная по некоторой наблюдаемой выборке данных выборочная функция плотности распределения (ВФПР).

В применяемых в настоящее время на практике моделях анализа и прогнозирования динамики нестационарных временных рядов (их эволюции) в качестве аппроксимаций выборочных распределений чаще всего используются диффузионные уравнения, включая нелинейную диффузию [30], уравнение Лиувилля [29], уравнение Фоккера – Планка [29] и ряд других.

Уравнение модели нелинейной диффузии имеет вид:

$$\frac{dp(x,t)}{dt} = D(t) \frac{d^2 p^{\frac{n-1}{n+1}}(x,t)}{dx^2},$$

где  $n$  – числовой параметр модели, а  $\frac{n-1}{n+1}$  – показатель степени функции плотности распределения  $p(x,t)$ . Это уравнение учитывает только случайные переходы.

Уравнение Фоккера – Планка имеет вид:

$$\frac{dp(x,t)}{dt} = \frac{D(t)}{2} \frac{d^2 p(x,t)}{dx^2} - \frac{d(U(x,t)p(x,t))}{dx},$$

где  $U(x,t)$  – скорость «сноса»,  $D(t)$  – коэффициент диффузии, и учитывает не только случайное изменение (член уравнения  $\frac{d^2 p(x,t)}{dx^2}$ ), но и упорядоченные нестационарные переходы (член уравнения  $\frac{d(U(x,t)p(x,t))}{dx}$ ) или «нестационарный снос».

Уравнение Лиувилля:

$$\frac{dp(x,t)}{dt} + \frac{d(U(x,t)p(x,t))}{dx} = 0$$

определяет упорядоченный перенос.

Практика показывает, что на самом деле экономические процессы не являются полностью случайным, а имеют тенденцию к самоорганизации, а также на них накладывае-ся возможность влияния памяти о предыдущих состояниях.

### 3. Моделирование процессов в социотехнических системах (социальные сети) на основе анализа временных рядов

При моделировании процессов в социотехнических системах (например, в социальных сетях) необходим математический аппарат, который позволил бы формализовать характер

данных и привести их к единой шкале измерений. Очевидно, что нельзя в одной модели проводить вычислительные операции, например, над лингвистическими оценками и величинами метрической шкалы без использования процедур отображения на формальное безразмерное множество. Эта проблема имеет исключительную важность, поскольку большинство данных, например, в социальных сетях имеют текстовый формат, и для получения временного ряда, описывающего динамику наблюдаемых процессов с течением времени, необходим их анализ.

В последнее время предпринимается множество попыток создания различных моделей и методов прогнозирования социальных, экономических и политических событий путем анализа текстовой информации и построения на её основе временных рядов. Рассмотрим некоторые методы, которые применяются для анализа плохо структурированной текстовой информации и прогнозирования событий.

В работе [31] для решения задачи анализа авторы изучают временные зависимости в потоках событий и вводят кусочно-постоянную аппроксимацию их интенсивности, применяя Байесовский подход и распределение Пуассона к описанию выборки важности будущих событий. Это позволяет построить нелинейные временные зависимости для предсказания будущих событий с использованием деревьев решений. Однако авторы не рассматривают наличие памяти информационного пространства и возможность его самоорганизации, что является, на наш взгляд, очень важным для прогнозирования редких событий.

В статье [32] на основе анализа текстовых данных и методов машинного обучения авторы разрабатывают модель для прогнозирования катастроф и стихийных бедствий. Они предлагают анализировать исторические данные, извлекать из них шаблоны событий, связанных с различного рода катастрофами и использовать эти шаблоны для машинного обучения в качестве обучающей выборки, а затем прогнозировать будущие катастрофы, используя текущие события. Авторы собрали данные о катастрофах из поисковой системы Google, делая запросы по ключевым словам. Затем текстовые документы, полученные в результате запросов, обрабатываются методами математической лингвистики, и с использованием обученного байесовского классификатора отсеиваются ложные результаты. После сбора данных авторы выполняют их смысловую кластеризацию. Из ключевых слов, по которым генерировались поисковые запросы, строится матрица переходов, а из сгруппированных событий строится матрица наблюдений. Затем обе матрицы подаются на вход скрытой марковской модели для составления прогноза. По мнению авторов, данный подход позволяет предсказывать будущие события и места, в которых эти события произойдут.

В работе [33] авторы также описывают модель прогнозирования будущих событий, путем обобщения конкретных наборов последовательностей произошедших событий, извлечённых из новостей за 22 года с 1986 по 2008 годы. Авторы предполагают, что события в реальном мире генерируются вероятностной моделью, которая так же генерирует новостные сообщения об этих событиях. Сообщения из новостных событий используются для построения модели в форме вероятности  $P(ev_j(t + \Delta) | ev_i(t))$  реализации некоторого будущего события  $ev_j$  в момент времени  $t + \Delta$  и произошедшего события  $ev_i(t)$  в момент времени  $t$ . Например, модель показывает, что с вероятностью 18% событие о засухе  $ev_j$  происходит после события о потопе  $ev_i$ . Эта вероятность аппроксимирует связь между двумя произошедшими событиями реального мира.

В работе [34] было исследовано влияние макро-новостей на экономические процессы, в частности на скачки валюты и цен. Используя макроэкономические новости, показывающие конкретные индексы, такие как уровень безработицы в стране, уровень инфляции и т.д., авторы доказали, что 9–15% скачков валюты напрямую связаны с новостями США, что новости об экономике США оказывают негативное влияние на скачки валюты и цен.

Хуанг (Huang) и соавторы в [35] рассмотрели эту проблему, разработав заранее подготовленную модель для информирования инвесторов о важных политических и экономических новостях в реальном времени. Модель основана на взвешенных правилах ассоциации, которые используются для определения того, является ли выпущенная новость достаточно важной для инвесторов. Во время обучения по реальным данным алгоритм взвешенных ассоциативных правил обнаруживает множественные термины, которые часто появляются одновременно в одном заголовке новостей. Интуитивное значение правила ассоциации заключается в том, что заголовки новостей в базе данных, которые содержат набор ключевых слов  $X$ , также содержат набор ключевых слов  $Y$ . Вес  $wks_p$  для отдельного ключевого слова определяется как

$$wks_p = \frac{\sum_j pf_{p,j}}{n},$$

где  $pf_{pj}$  обозначает колебание цены закрытия акции на следующий торговый день, и ключевое слово  $p$  появляется в заголовке (-ах) новостей на  $j$ -ый день,  $n$  представляет собой общее количество дней, в которых ключевое слово появляется в заголовках новостей. Таким образом, эти данные помогают решить, влияют ли ключевые слова в заголовках новостей на результат торговли. Модель была откалибрована на реальных данных путем вычисления правил ассоциации весов и построения базы данных весов ключевых слов. Экспериментальные результаты показывают ее выполнимость для анализа новостей в режиме реального времени.

Робертсон (Robertson) и его коллеги исследовали влияние на рынок не-макроэкономических новостей, а также времени суток и процента новостей, когда рынок реагирует наиболее значимо [36]. Они определили гетерогенные временные ряды средней цены акций, объема, количества сделок и собрали данные по каждой минуте в рамках бизнес-дня (от начала до конца торгового периода). Затем был определен временной ряд доходности, использованный для определения периодов высокой доходности, свидетельствующих о реакции рынка на новости. Кроме того, изменение временных рядов объема и изменение временных рядов количества торгов использовались для обнаружения внезапного увеличения объема и количества торгов в качестве индикаторов реакции рынка на новости.

В работе [37] авторы разработали алгоритм обнаружения причинно-следственных связей между событиями в социальных сетях и применяют его для прогнозирования точности и времени реализации событий. Алгоритм основан на поиске и анализе коротких сообщений в Твиттере. Твиттер – это социальная сеть для публичного обмена сообщениями при помощи веб-интерфейса, SMS, средств мгновенного обмена сообщениями или сторонних программ-клиентов для пользователей интернета любого возраста. Пользователи Твиттера публикуют короткие сообщения – твиты, которые могут содержать

фотографии, видео, ссылки и текст. Эти сообщения появляются в профиле пользователя и доступны всем друзьям автора сообщения в социальной сети. Разработанный алгоритм состоит из нескольких шагов. На первом шаге отбираются твиты за определенный период времени, из твитов извлекаются ключевые слова. На втором шаге по ключевым словам определяется тональность извлечённых слов – положительная, отрицательная или нейтральная. Для определения тональности слов используется обученный классификатор на базе метода опорных векторов (SVM). На третьем шаге определяются причинно-следственные связи между ключевыми словами, для этого используется методика обучения ассоциативным правилам (association rule learning), которая извлекает правила «if-then» из данных. На четвертом шаге выполняется прогнозирование событий, на основе результатов временного анализа твитов и причинно-следственных связей, вычисленных на предыдущем шаге.

В работе [38] представлены исследования, посвященные использованию пространственно помеченных твитов, для прогнозирования преступности в масштабе крупного города. Основная гипотеза заключается в том, что местоположение, время и содержание твитов информативны в отношении прогнозирования времени и места появления возможных криминальных событий.

Традиционными методами анализа и визуализации распределения преступлений в городском пространстве и времени являются метод карты горячих точек, вычисляющий двумерную пространственную функцию плотности вероятности для исторического преступления и метод моделирования самовозбуждающихся точечных процессов, фиксирующий пространственно-временную кластеризацию криминальных событий. Ограничениями указанных методов при анализе преступности являются: локальная описательность (неприменимость к различным географическим регионам), требование наличия исторических данных о преступности для интересующего региона, отсутствие учета показателей социального медиа-ландшафт региона.

Для выделения тематик твитов автор предлагает использовать тематическое моделирование, т.е. построение модели коллекции текстовых документов, определяющей к каким темам относится каждый документ коллекции (в данном случае твит). Алгоритм построения тематической модели получает на входе коллекцию текстовых документов. На выходе для каждого документа выдаётся числовой вектор, составленный из оценок степени принадлежности данного документа каждой из тем. Размерность этого вектора, равная числу тем, может либо задаваться на входе, либо определяться моделью автоматически.

Затем проводится токенизация – замена конфиденциального элемента данных на неконфиденциальный эквивалент, называемый токеном, не имеющий самостоятельного смысла/значения для внешнего или внутреннего использования. Токен – это ссылка (т.е. идентификатор), сопоставимый с конфиденциальными данными через систему токенизации. Система токенизации предоставляет приложениям обработки данных полномочия и интерфейсы для запроса токенов или расшифровку конфиденциальных данных из токенов.

Для определения правильной части речи для слова из твита (тегирование части речи) использовался частичный речевой теггер. Его задача – автоматическая обработка текста для определения части речи и грамматических характеристик слов в тексте

(корпусе) с приписыванием им соответствующих тегов. Важную роль в анализе имеют смайлы (стилизованные графические изображения человеческих эмоций). Созданный автором специальный токенизатор распознает смайлы как отдельные токены, несущие в случае твитов семантический контент, описывающий эмоциональное состояние пользователя, и использующиеся для моделирования тематики твита.

Для проверки результатов прогнозирования автор выделил квадрат на карте города и в течение месяца соотносил предсказанные преступления с реальными данными. На этой выборке был обучен бинарный классификатор. Результаты показали, что использование методов лингвистического анализа и статистического моделирования для автоматического определения популярных городских тем на основе данных из твитов городских жителей и включение собранного списка обсуждаемых тем в модель прогнозирования преступности улучшают эффективность стандартных методов прогнозирования преступности в несколько десятков раз.

Подводя итог, необходимо отметить, что переписка в социальных сетях, например в сети Twitter, является идеальным источником данных для поддержки принятия решений. Пользователи Twitter, исчисляемые миллионами, публично и эмоционально обсуждают актуальные события и темы, создавая контент, бесплатно распространяемый по всему миру в режиме реального времени, в том числе и в виде сообщений (твитов), помеченных точными пространственными и временными координатами. Также выявлено, что одной из главных задач при моделировании процессов в социотехнических системах (например, социальных сетях) является разработка математического аппарата для приведения данных к единой шкале измерений.

#### **4. Моделирование процессов в социальных системах (электоральные кампании) на основе анализа временных рядов**

Рассмотрим электоральную кампанию как сложную систему, некую структуру взаимосвязанных объектов, в которых присутствует человеческий фактор, уже в силу своей природы приводящий к нечеткости, неопределенности и недетерминированности характеристик, наблюдаемых в сложных структурах электоральных процессов.

Сложность систем определяется свойством эмерджентности. Возникновение эмерджентности не является итогом простого суммирования показателей характеристик элементов системы, а проявляется в результате возникновения цепочек системных связей и адаптивного перераспределения функций между элементами. Одним из проявлений эмерджентности, например, является самоорганизация и память о предыдущих состояниях, связанная с наличием человеческого фактора [39].

Описывающая электоральный процесс модель должна подразумевать решение задачи прогнозирования. Однако при изучении электоральных процессов не всегда возможно описать изучаемый процесс математически, даже при наличии данных по некоторой характерной наблюдаемой величине. В таких случаях используется подход, основанный на анализе сигналов, производимых сложной системой. При упоминании таких сигналов принято использовать модели на основе временных рядов.

Прогнозирование результатов выборов – относительно недавняя и все более популярная часть социологических и политологических исследований. Конкуренционные выборы

являются отличительной чертой современной демократии, и прогнозирование победителя интересует всех участвующих или наблюдающих за ходом избирательной кампании.

Прогнозирование выборов выделяется из других политологических исследований, так как прогноз:

- основан на статистических и оперативных данных, которые характеризуются ограниченностью и недостоверностью;

- ориентирован на временно-ограниченную задачу;

- связан с погрешностью методов измерения исходных данных;

- связан со спецификой проведения избирательной кампании и непосредственно выборов.

Во время проведения избирательной кампании и выборов собирается информация об экономических, социальных, политических условиях жизни электората; выявляется состав основных электоральных слоев и групп, в рамках которых определяются актуальные проблемы, политические и партийные предпочтения избирателей.

Уже в процессе избирательной кампании начинается анализ собранных данных, ищутся тренды дальнейшего развития выборного процесса, делаются попытки управления ходом избирательной кампании. После того как результат выборов становится известным, идет подведение итогов выборов и выявление причин, приведших к успеху или неудаче кандидатов.

Несмотря на то, что во время проведения избирательной кампании и последующего анализа результатов выборов собирается большой объем разнотипных данных, требующих аналитической обработки, активное применение математических методов и основанных на них компьютерных технологий для прогнозирования или оценки результатов выборов до сих пор является в среде социологов и политологов достаточно спорным вопросом. В ходе исследования ВЦИОМ по изучению американских практик электорального прогнозирования с использованием данных опросов общественного мнения [40] были проведены интервью с ведущими специалистами в области электоральных исследований и прогнозирования, в которых участвовали представители Pew Research Center, CBS News, YouGov, Edison Research, SurveyMonkey, Gallup, Wilson Perkins, Allen Opinion Research, Huffington Post, Abt SRBI, SSRS, Langer research associates, Института социально-экономических исследований и политики (ISERP), исследователями Колумбийского, Принстонского и Стэнфордского университетов, отделов поллинга телекомпаний ABC News и CBS News, отдела опросов The Washington Post, а также ряд независимых исследователей, занимающихся электоральным прогнозированием. Результаты интервью показали, что среди американских исследователей электоральных процессов можно выделить четыре группы по типу решаемых задач:

- «публичные» социологи (решение повседневных исследовательских задач широкого профиля),

- партийные социологи (решение социологических задач с учетом партийных интересов),

- академические социологи (работа с агрегированной информацией по выявлению зависимости результата выборов от набора различных индикаторов),

- агрегаторы (исключительно агрегирование социологических данных).

По мнению опрошенных, аналитические модели академических исследователей характерны тем, что рассматривают большое количество индикаторов различной природы,

которые позволяют получать структурные оценки вероятных результатов выборов на основе статистических данных предыдущих выборов [40]. При этом единственным показателем, который получен из текущих опросов общественного мнения, является рейтинг одобрения деятельности настоящего избранника: «по существу берется самый новый рейтинг популярности Президента и прогоняется через статистические модели с временными рядами» [40]. Поэтому полстеры (от английского pollster) – специалисты, которые занимаются подготовкой, проведением и анализом результатов опросов общественного мнения (polls) – редко используют академические модели, считая их больше ретроспективными и «объяснительными», и мало учитывающими текущие настроения электората и специфические условия проведения выборов в том или ином государстве.

Часть российских социологов и политологов разделяют мнение о том, что формализованное математическое описание сложных разнородных стохастических электоральных процессов приводит к сильной примитивизации модели выборного процесса и не дает устойчивого достоверного результата. В Общественной Палате Российской Федерации (ОПРФ) состоялся круглый стол «Математика на службе избирателей», в ходе которого обсуждалась значимость математических исследований в оценке избирательного процесса [41]. Основное внимание было уделено возможности доказательства фальсификации результатов выборов с помощью математических методов (теория Шпилькина). Заместитель Секретаря ОПРФ, руководитель рабочей группы ОПРФ по мониторингу реализации избирательных прав граждан Александр Точенов отметил: «Некие математические методы накладываются на социальные процессы, и мы не получаем подтверждения того, что данные объективны. На мой взгляд, математика в избирательном процессе стала партийной, потому что социальные системы неравномерны, нелинейны, неоднородны. Цифры — это влияние на сознание, а избирательная система должна быть безупречной и достоверной» [41].

Однако, автор работ [42, 43], отмечает, что такие проблемы исследований, как сопряжение заложенной в математическом методе модели с содержанием решаемой социологической задачи и невозможность проверки репрезентативности социологической выборки, относятся скорее к самой методике социологического исследования, и на него не оказывает влияния применение или неприменение математических методов и компьютерных технологий. Далее автор в своих статьях описывает перспективные для социологических исследований математические методы и компьютерные технологии, показывая полезность их применения в данной предметной области. Например, в [42] показана роль методов математического моделирования социальных процессов, масштаб развития которых, по мнению автора, позволяет говорить о них как об относительно автономной ветви науки. В статье [43] тот же автор отмечает, что «для обеспечения эффективной согласованности теоретического и эмпирического этапов исследования требуется, чтобы процесс применения математического аппарата сводился к построению серии таких математических моделей социальной реальности, которые отражали бы теоретические представления социолога об изучаемом явлении. Этому могут способствовать: а) понимание математики не только как строгой формальной науки (традиционная точка зрения), но и как научной ветви, направленной на выработку правил построения модели изучаемой реальности, записанной с помощью математического языка; б) рассмотрение

всех моделей, лежащих в основе математических методов, как результата обобщенного измерения и обеспечение соответствия этих моделей априорным теоретическим представлениям; в) использование определенных методологических принципов применения математического аппарата в социологии». Таким образом, автор делает вывод, что потребности политических и социологических наук в изучении количественной стороны электоральных процессов, а также развитие методов и средств информационных технологий неизбежно приведут к математизации политических и социологических исследований электоральных процессов.

В [40] также отмечается, что, несмотря на специфичность проведения избирательной кампании и выборов в различных государствах, например, в Российской Федерации количество избирателей на последнем этапе выборов существенно больше, чем в США, а количество факторов, влияющих на предпочтения избирателей в России, существенно меньше, чем в Америке. Также можно выделить набор задач, которые являются общими для электоральных прогнозов: поддержание публичности прогноза (т.е. прогноз должен быть опубликован); жёсткое соблюдение методических требований социологических опросов (тип выборки, сроки проведения опроса, процедуры контроля, методика опроса). Общемировой тенденцией является расширение применения новых ИТ-технологий при проведении опросов, создании соответствующих методик сбора и обработки данных, прогнозирования результатов. Поэтому разработка новых математических методов и моделей прогнозирования результата выборов является востребованной задачей.

#### **4.1. Структурные модели**

Первые электоральные прогнозы были сделаны в компании Gallup при попытке спрогнозировать президентские выборы в США в 30–40-х годах, потом в течение сорока лет политические партии и отдельные научные группы предлагали прогнозы выборов на основе данных, полученных в ходе собственных предвыборных опросов. В 80-ые годы появились первые модели, использующие регрессионные подходы с учетом экономических и политических индикаторов, предположительно влияющих на популярность кандидатов. В то время наиболее популярным математическим подходом был метод наименьших квадратов, основанный на минимизации суммы квадратов отклонений некоторых функций от искомым переменных [44]. Метод наименьших квадратов является одним из базовых методов регрессионного анализа для оценки неизвестных параметров регрессионных моделей по выборочным данным.

В России одним из первых ученых, работавших в данной области, был А.О. Крыштановский (1955–2005 гг.) – советский и российский социолог, математик, специалист по методикам социологических исследований и анализа социологических данных, первый декан факультета социологии Высшей школы экономики. За 1995–2005 гг. Крыштановским А.О. были опубликованы работы [45–48], посвященные особенностям и возможностям применения математического аппарата временных рядов для анализа и прогнозирования электоральных процессов. Автор отмечал, что применение математического аппарата временных рядов позволит изменить единицу анализа динамики электоральных процессов, перейдя от анализа характеристик отдельного респондента к анализу социальной группы [45, 46]. Был рассмотрен ряд вопросов, возникающих при смене едини-

цы анализа с отдельного респондента на социальную группу. Первый – это построение индикаторов, характеризующих новую единицу анализа – социальную группу. Если рассматривать, например, средний возраст социальной группы, то данный индикатор можно получить простым усреднением возрастов респондентов. В других случаях задача не является настолько тривиальной и требует отдельного рассмотрения [45]. Вторым вопросом – наличие неодинаковых интервалов времени между опросами. В приведенных исследованиях интервал составлял один-два месяца, поэтому нарушение требования равенства временных интервалов было устранено путем удаления части результатов исследований в данных мониторинга, а интервал установлен равным двум месяцам. Для исследований была использована традиционная формула представления временного ряда, где значение анализируемого показателя считается как сумма значения тренда (плавно меняющегося, нециклического компонента, описывающего чистое влияние долговременных факторов, эффект которых сказывается постепенно), сезонного компонента, описывающего регулярные изменения в течении заданного периода (года, месяца, недели и т.п., последовательность циклов предполагается почти точно повторяющейся), циклического компонента, описывающего длительные периоды подъемов и спадов (сами циклы могут меняться по протяженности и амплитуде) и ошибки (отклонения реальных данных от суммы первых трех закономерных составляющих). Далее автор, используя методы регрессионного анализа, рассматривал выделение обобщенных показателей для каждого из опросов, например, показатель среднедушевого дохода.

Крыштановский А.О., как и большинство авторов структурных моделей электорального прогнозирования, ставил результат выборов в линейную зависимость от результатов опросов и влияния экономических (безработица, инфляция, рост национального дохода) или внеэкономических (популярность кандидата, степень партийной поддержки претендента, роль личного фактора) индикаторов с учетом погрешности рассматриваемых показателей. Разумеется, что в различных политических системах на активность избирателей влияют самые различные факторы. В США, например, уровень активности избирателей, более всего зависит от уровня образования, пола, возраста, политического интереса. Например, в [49] описано изменение политических предпочтений (поляризация взглядов) врачей в Соединенных Штатах во время кампаний на федеральных выборах с 1991 по 2012 год и их финансовых вкладов в фонды республиканцев и демократов в зависимости от пола и врачебной специализации. В частности, общий объем взносов увеличился с \$20 млн. (в 1991 году) до \$189 млн. (2012 год), а доля врачей, делающих взносы, – с 2.6% до 9.4%. Взносы в фонд республиканцев были более распространены среди мужчин, чем среди женщин (52.3% против 23.6%). Врачи, практикующие в государственном секторе, составили 53,2% от всех делавших взносы, а частном – 25.6%; хирурги – 70.2%, педиатры – 22.1% и т.д. Общий вывод, который был сделан авторами работы [45] на основании представленных ими данных, заключается в том, что в период с 1991 по 2012 год политические предпочтения американских врачей сместились от республиканцев в сторону демократов. Однако причины такого изменения проанализированы не были.

В работе [50] авторы описали динамику изменения предпочтений избирателей в Конгресс США на промежуточных выборах 2010 года. Исследования показали снижение поддержки кандидатов от демократов по отношению к республиканцам на завершающем

этапе кампании (отметим, что тоже самое произошло при выборе президента США в 2016 году – Трамп обошел Клинтон). На основе описанной динамики автор [50] сделал вывод, что данные изменения связаны с колебаниями в утверждениях Обамы, ухудшением национальных экономических условий и правительственной реформой здравоохранения.

В Италии и Австрии такие характеристики как образование, пол, возраст, политический интерес практически не влияют на выбор. В Англии, Германии, Швейцарии, Нидерландах на уровень активности также не оказывают влияния ни образование, ни пол [44].

В исследованиях российских специалистов большое место занимают вопросы политической культуры и ее влияния на активность избирателей. Например, в [51] авторы предложили следующий методологический подход к определению основных тенденций электоральных предпочтений российских избирателей. На первом этапе были выделены наиболее значимые параметры, на основе которых могут формироваться электоральные предпочтения. Это – место проживания (принадлежность к определенной электоральной географии), динамика экономического благосостояния, протестные настроения, категории населения (возраст/пол/трудовой статус и др.). К дополнительным значимым параметрам были отнесены: влияние региональных средств массовой информации (СМИ), динамика политических предпочтений (статистика ВЦИОМ), работа партий в регионах, влияние ошибок участковых избирательных комиссий (УИК) на достоверность информации, влияние административного ресурса, вид выборов (регион/местные/федеральные и партии/персоналии).

Затем для проверки адекватности предположения о влиянии различных индикаторов на модели был выполнен анализ статистических данных протоколов УИК. Поскольку данными социологических опросов авторы не располагали, то были взяты данные протоколов УИК 25 регионов европейской части РФ по выборам Президента РФ 2 марта 2008 года. На территории этих регионов численность электората составила 22 223 670 человек, что составляет около 21% избирателей страны, количество УИК – 22 318 (примерно 23% УИК). Был сделан вывод о верности гипотезы о неоднородности множества избирателей. Подтверждена независимость распределений по независимым параметрам (например, зависимость количества УИК от явки избирателей) на однородных подмножествах избирателей. Отмечена необходимость сравнения расчетных данных по модели с реальными данными электоральной статистики.

Структурные модели отличаются друг от друга, в основном, оценками политических и экономических показателей. Например, прогнозируя лидера на президентских выборах во Франции в 2012 году, авторы [52] оценивали функцию голосования на основе данных о местных и национальных выборах с 1981 по 2007 годы и политических факторах. Окончательное прогнозирование победителя на данном политическом фоне определялось отслеживанием изменения экономических факторов (безработица). В результате был дан прогноз на поражение Н. Саркози.

К экономическим факторам также часто относят изменения в ВВП на душу населения, инфляцию и безработицу, рост реальных доходов и субъективные убеждения избирателей в экономических показателях правительства. Зависимость прогноза победителя от изменения этих факторов считалась линейной. Например, в [53] показано, что на конечный выбор избирателя влияют не совокупные успехи экономической политики за

весь срок правления, а результаты последнего года перед выборами, что скорее относится к области психологии, а не социологии и политологии.

В работе [54] для описания изменения предпочтений избирателей во время выборов президента в Ирландии в 1997 году использовалась модель смешанного членства в разных группах. Согласно этой модели один и тот же человек может быть одновременно отнесен к группам с разными взглядами, т.е. не имеет четко выраженных предпочтений.

Коэффициенты, используемые в структурных моделях при оценке экономических и политических переменных, генерируются путем применения прогнозной модели к как можно большему количеству прошедших выборов, чтобы узнать, какое влияние эти переменные оказали в прошлом. Таким образом, прогнозирование следующих результатов обеспечивается добавлением соответствующих значений коэффициентов для политических и экономических показателей в том виде, в каком они находятся в год выборов, а затем умножением их на коэффициенты, которые, как говорят нам предыдущие выборы, обеспечивают наилучшее соответствие.

Впервые модель влияния экономики на результаты выборов разработал Рэй Фэйр (Ray Fair), профессор экономики Йельского Университета (Yale University), в 1980-е годы [55]. Были выделены два базовых фактора, влияющие на результат выборов: темпы роста или снижения валового внутреннего продукта США в течение трех кварталов перед днем голосования (выборы в США традиционно проводятся в первый вторник ноября) и средний уровень инфляции на протяжении предыдущих 15-ти кварталов. Модель Фэйра позволила точно предсказать результаты пяти из прошедших семи президентских выборов – 1980, 1984, 1988, 2000 и 2004 годов.

Известный политолог Даглас Хиббс (Douglas Hibbs) в 2000 году предложил иную модель, назвав ее «Хлеб и Мир» (Bread and Peace). Базовыми показателями модели были выбраны:

– среднестатистический темп роста/снижения чистого дохода на душу населения в течение четырех лет, предшествующих президентским выборам (президентская каденция длится четыре года), т.е. повышение реального дохода на 0.01% приводит к увеличению числа голосов избирателей, отданных за правящую партию, на 0.36% (этот прирост возможен только в том случае, если рейтинг правящей партии составляет не менее 46%);

– количество американских солдат, погибших в иностранных войнах (особенно в Корее и Вьетнаме). По расчетам Хиббса, каждая тысяча убитых американских солдат отбирает у правящей партии 0.3% голосов.

Модель Хиббса была подтверждена результатами 12 из 14 выборов, прошедших в период с 1996 по 2000 год. В 2004 году она также позволила верно определить победителя.

Автор работы [49] утверждает, что эти две вышеописанные структурные модели имеют преимущество перед опросным прогнозированием, т.к. информация, необходимая для прогноза, доступна до следующих выборов, и сами модели основываются на оценке данных, которые действительно важны для избирателей, иначе прогноз будет неверным.

Существенным фактором, препятствующим возможности прогнозирования исхода избирательных кампаний, является большая волатильность предпочтений избирателей. В работе [56] для возможного прогнозирования итогов голосований и снижения волатильности предлагается методика использования в опросах вопросов типа «скорее все-

го». По мнению авторов работы [56], значительная часть изменения социологических данных о предпочтениях избирателями кандидатов определяется не реальными сдвигами избирателей в предпочтении, а скорее всего изменениями в составе вероятного пула избирателей. А вопросы типа «скорее всего» могут позволить разделить реальные изменения предпочтений избирателей и изменения их пулов (изменения мнения данного избирателя, от его замены в выборке на другого человека, т.к. исследования являются обезличенными).

Волатильность предпочтений избирателей и влияние политической искушенности на их окончательное электоральное решение во время региональных выборов в Бельгии в 2009 году было изучено в работе [57]. Рассмотрены причины и мотивы изменения предпочтений. Результаты показывают, что избиратели с низким уровнем политической эффективности (понимания программ кандидатов и партий и интересом к ним), как правило, чаще меняют свои предпочтения, как в рамках одной кампании, так и между выборами. Избиратели с более высоким уровнем интереса к политике переключили свои намерения до начала кампании. Это в значительной степени определяет волатильность.

Обзор моделей выбора предпочтений при голосовании и избирательной политики представлен в работе [58]. В ней обсуждаются два класса моделей: те, в которых выборы рассматриваются с позиции агрегации предпочтений, и те, которые смотрят на выборы, исходя из механизмов агрегации информации. Модели первого класса, как правило, учитывают общее изменение с течением времени доли избирателей (являются феноменологическими, не связывают изменения и вызывающие их причины) с определенными предпочтениями, а модели второго класса учитывают влияние общего количества информации о кандидатах или партиях на избирателей во время выборов (частично связывают изменения с вызывающими их причинами). Кроме того, по мнению авторов работы, необходимо учитывать роль асимметрии информации (политики лучше информированы о политике, чем избиратели, что позволяет им предпринимать скрытые действия).

Необходимо отметить, что существенную роль на предпочтения избирателей оказывает информационное пространство (телевидение, пресса, интернет, социальные сети и т.д.). В частности, в работе [59] рассматривается роль воздействия на предпочтения избирателей средств массовой информации в бразильских президентских кампаниях 2002 и 2006 годов. В этой работе учитывались следующие факторы: общая партийная пропаганда кандидатов, время свободных публичных выступлений кандидатов, освещение их деятельности комментаторами. В своей работе P.S. Mundim делает вывод, что партийная пропаганда и выступления самих кандидатов оказало большее воздействие на результаты выборов в Бразилии, чем освещение их деятельности комментаторами.

## **4.2. Опросные модели**

Опросы общественного мнения уже давно играют важную роль в изучении и проведении выборов. В работе [60] подробно описана история проведения и использования результатов опросов для прогнозирования результатов выборов, понимания поведения избирателей и планирования стратегии избирательной кампании. Автор рассматривает опросы как базовый инструмент, который ученые и журналисты используют для понимания поведения избирателей. В последнее десятилетие отмечается, с одной стороны, ко-

личественный рост опросов и изменение средств проведения опросов: активное использование интернет-опросов, опросов при помощи мобильных телефонов, интерактивных голосовых ответов (IVR); с другой стороны, отмечено снижение качества данных из-за методологических ошибок обработки собранной информации. Последнее затрудняет использование данных опросов при составлении избирательных прогнозов. Также автор прослеживает тенденцию дополнения данных опросов различными методами и моделями. Тем не менее, опросы являются неотъемлемой частью политических избирательных кампаний. Давний скептицизм в отношении того, что респонденты могут сказать что-либо о мнениях миллионов, уступил место вере в научную основу вероятностных выборок. Неперспективные выборки – как правило, выборочные интернет-опросы – становятся все более распространенными, а вероятностные выборки сталкиваются с серьезными методологическими проблемами, такими как увеличение числа респондентов, нежелающих принять участие в опросе или имеющих только мобильные телефоны.

Информационные технологии предложили новые инструменты проведения опросов. Сегодня опросы общественного мнения распространены не только в преддверии выборов, но и на протяжении всего избирательного цикла, охватывают большее число респондентов. Однако вопрос достоверности полученных данных, корректности агрегирования данных остается актуальным.

В работе [61] отмечается, что в 2016 году социологи накануне общенациональных президентских выборов провели более 400 опросов, в среднем около двух в день. Каждый выпуск опросов пристально изучался и комментировался представителями средств массовой информации. Однако, как считает автор статьи, несмотря на сфокусированность внимания прессы на результатах опросов и увеличение количества опросов, сама модель политического опроса осталась той же, что была до появления интернета. Главный фактор достоверности результата – это методика проведения опроса. Автор отмечает, что до широкого использования идентификатора вызывающего абонента телефонные опросы проводились намного проще. В 1997 году для национальной выборки из примерно 800 респондентов требовалось от 2000 до 2500 звонков. Сегодня, чтобы получить такое же количество респондентов, требуется от 7500 до 9000 звонков, чтобы получить образец разумного размера, что является резким снижением доли ответивших. Это приводит к тому, что выборка фактически не отражает группу, которую пытается измерить опрос.

Автор [61] отмечает удивительность того факта, что для того, чтобы оценить предпочтения населения страны, достаточно опросить всего 800 людей. Но этот факт подтверждается, если 800 человек действительно выбраны случайным образом. В качестве подтверждающего примера рассмотрена ситуация после провальных для Б. Обамы его первых дебатов против М. Ромни в 2012 году, когда избиратели-демократы просто стали реже отвечать на опросы, поскольку они просто не хотели говорить о политике, в то время как республиканцы повысили свою активность. В итоге результаты опроса сместились в сторону М. Ромни, хотя его реальная поддержка не увеличилась: просто выборки были предвзяты по отношению к избирателям М. Ромни.

Для достижения случайности выборки при наиболее достоверных телефонных опросах опросчикам приходится увеличить количество звонков до семи или более попыток на одного человека. Несмотря на эти усилия по получению случайной выборки, показатели

отклика остаются низкими среди представителей молодежи, говорящих на испанском языке, евангелистов и афроамериканцев [61]. Чтобы исправить этот тип смещения результатов опросов, опросчики используют взвешивание. Например, взвешивание выполняют тогда, когда 6% выборки составляют афроамериканцы, а доля афроамериканцев составляет 12% населения США. Автор считает, что применение процедуры взвешивания на основе известных демографических факторов (раса, возраст и пол) – это уже искусство, а не наука, так как социологи не могут быть уверены, что члены группы, к которой обратились, представляют группу в целом, и их ответы поддержат недостающие проценты участников группы. Например, большинство политических опросов проводится только на английском языке, чтобы избежать острых проблем перевода и более высоких зарплат двуязычных интервьюеров, но 73% латиноамериканцев говорят на испанском языке дома, поэтому выборка латиноамериканцев, достигнутая большинством опросов, не репрезентативна для всех латиноамериканцев.

Также нет полного согласия относительно того, можно ли взвешивать политические факторы: если в опросе доля демократов или республиканцев невелика, следует ли использовать взвешивание, чтобы исправить это? Подобные решения дают некоторым опросчикам возможность сдвигать свои результаты в ту или иную сторону, чтобы избежать слишком большого отклонения от результатов других опросов. Поскольку средние результаты опросов стали более доступными и распространёнными, некоторые исследователи не публикуют свои результаты опросов, если их результаты отличны от средних значений опросов, либо используют стратегическое взвешивание, чтобы вернуть свои опросные результаты к средним значениям опросов. В итоге таких манипуляций исследователи могут пропустить сдвиги в общественном мнении, выявляемые опросами. Лучшие телефонные интервьюеры имеют большой опыт работы и высшее образование, и их оплата – главная стоимость политических опросов. Поэтому попытки автоматизировать этот процесс, чтобы сэкономить деньги, используя интерактивные голосовые ответы (IVR), при котором интервьюеры в режиме реального времени заменяются записанными подсказками, а респонденты дают ответы, разговаривая с компьютером, вполне понятны. Эти услуги значительно ускоряют и удешевляют опрос: они могут заполнить его за несколько часов и стоить примерно 1/10 от стоимости живых интервьюеров. Однако, IVR-опросы имеют более низкие показатели отклика, чем традиционные выборки телефонов, способствуют большему количеству ложных ответов, не могут легально связаться с сотовыми телефонами. Также есть категории избирателей, при опросе которых использование IVR-опросов бесполезно.

Онлайн-опросы еще более дешевые и быстрые по сравнению с традиционными телефонными опросами, но 16% американцев не пользуются интернетом, что требует дополнительного уровня взвешивания, чтобы попытаться приблизиться к репрезентативной выборке общественности [61]. Чтобы обойти этот факт приходится предварительно звонить и договариваться с респондентом о том, что он пройдет онлайн-опрос. Это увеличивает стоимость и требует применения еще более сложного взвешивания, чтобы исправить тот факт, что результаты опросы повторно отбираются из результатов других опросов. Недавние исследования показывают, что даже после взвешивания онлайн-опросы, как правило, преувеличивают долю мужчин и безработных, возможно, потому, что

онлайн-опросы обычно требуют от людей выбора, а у безработных мужчин может быть больше мотивации делиться своими взглядами и с большей вероятностью иметь время для опроса. Это может помочь объяснить, почему Дональд Трамп, кажется, добивается большего успеха в онлайн-опросах, чем в телефонных опросах. Таким образом, традиционные телефонные опросы в режиме реального времени, по мнению автора [61], дают более точные результаты, чем альтернативные варианты.

Много исследований посвящено повышению достоверности агрегирования данных опросов. В статье [62] на основе опросов, полученных во время президентских выборов в США до 2016 года, была выполнена оценка эффективности агрегирования опросов с основных веб-сайтов с прогнозом. Было установлено, что незначительная погрешность данных, возникшая из-за поправок, внесенных в данные по партийным предпочтениям, оказала существенное влияние на результат прогноза на уровне штата. Авторами предложена новая модель сглаживания подобных эффектов, показавшая аналогичные или превосходящие результаты по сравнению с другими методами агрегации на нескольких выборах. Улучшение предложенной модели по сравнению с конкурирующими методами было особенно значительным в 2016 году из-за колебаний избирателей в поддержке кандидатов. Авторы [58] считают, что усиление акцента на фундаментальных статистических компромиссах смещения и дисперсии поможет в дальнейшем улучшить прогноз.

В статье [63] отмечается, что авторы за пять последних президентских выборов в США публиковали результаты прогнозирования, базирующихся на двух параметрах. Первый – результат опросов, второй – средневзвешенное значение квартального роста в LEI (Индекс ведущих экономических показателей), где каждое текущее квартальное значение взвешивается в 0.80 раза больше, чем в следующем квартале. Оценивался совокупный взвешенный рост ведущих показателей в тринадцатом квартале избирательного цикла, первом квартале года выборов (далее – совокупный рост LEI). Измеренный в 13 квартале совокупный рост LEI отражает как экономические изменения, произошедшие в течение избирательного цикла, так и ожидаемые – на оставшуюся часть года выборов. Совокупный рост LEI является опережающим экономическим показателем и, по мнению авторов, хорошо отражает экономические тенденции по мере приближения выборов и может быть измерен заранее. Соответственно, прогнозная модель авторов, основанная на опросах, проведенных в пробном периоде, и совокупном росте ведущих экономических показателей, предсказала в 2016 году победу Х. Клинтон над Д. Трампом.

В отчете Специального комитета по выборам в США 2016 года «Оценка выборов 2016 года в США», опубликованном Американской ассоциацией изучения общественного мнения [64], подробно разобраны причины несовпадения результатов опросов и прогнозов победителя президентской кампании с результатом выборов. Отмечено, что предварительные опросы отдавали предпочтение победе Х. Клинтон с вероятностью 90%, а оценки колебались от 71% до более 99%. В отчете представлены результаты анализа комитетом методов проведения и содержания предвыборных опросов в 2016 году, их сравнение с результатами предыдущих выборов и степень изменения результатов в зависимости от структуры опроса. Также были рассмотрены возможные причины расхождений, заявлены поздние колебания в предпочтениях избирателей по отношению к Д. Трампу и повсеместный сдвиг опросов к мнению выпускников колледжей, отдающих

свои голоса Х. Клинтон. Еще в качестве одной из причин рассматривают изменение явки избирателей, которая благоприятствовала победе Д. Трампа. В качестве системной ошибки при опросах отмечена неточность вероятной модели избирателей – основной причины систематической ошибки при опросе.

Модель, позволяющая просчитать победителя не среди двух участников, а предсказать победителя выборов в условиях многопартийной системы предложена в статье [44]. В качестве примера выбраны Германия (не меньше 6 партий) и Швеция (8 партий), где выборы традиционно являются многопартийными, и где модель прогнозирования результатов выбора, по мнению автора, не может быть чисто структурной моделью, т.е. учитывать большую долю голосов избирателей, как зависимую переменную, а экономические и политические показатели – как объяснительные переменные. Модель прогнозирования победителя многопартийных выборов должна включать данные опросов избирателей, и сделать это можно через динамическую линейную модель, принадлежащую общему семейству моделей пространственных состояний. Скрытая тенденция популярности поддержки оценивается путем объединения результатов опроса во временные ряды посредством фильтрации Калмана. Прогноз выборов затем делается путем экстраполяции этой тенденции на будущее.

В работе [65] на основе опросов общественного мнения в Великобритании представлены следующие методы прогнозирования мест в парламенте:

- корректировка и агрегирование опубликованных данных о намерениях избирателей по различным опросам;
- прогнозирование того, как общественное мнение может измениться до дня выборов;
- прогнозирование общего числа мест на основе прогнозируемых долей голосов в день выборов.

Авторами рассмотрена модель пространства-состояния для опросов общественного мнения, которая корректирует «домашние эффекты» и другие источники ошибок опроса, оценку исторических взаимосвязей между опросами и долей голосов в день выборов, а также вероятностный подход к прогнозированию победителя для каждого избирательного округа.

Автор статьи [66] представляет динамическую байесовскую модель, позволяющую прогнозировать результаты президентских выборов в США на государственном уровне. Предлагаемый метод на основе системного анализа объединяет информацию из исторических моделей прогнозирования в режиме реального времени с результатами большого количества опросов общественного мнения на уровне штатов, которые публикуются в ходе кампании. Результатом является набор прогнозов, точность которых изначально соответствует точности исторической модели, а затем постоянно улучшается на основе текущих данных. Автор на основе иерархической спецификации сглаживает временные и географические ограничения модели, параллельно используя и случайные тесты, подходящие по временным условиям. Таким образом, построенная модель отфильтровывает ежедневные колебания в опросах из-за ошибки выборки и эффектов национальной кампании, что позволяет ежедневно отслеживать предпочтения избирателей по отношению к кандидатам в президенты на государственном и национальном уровнях. Методы моделирования используются для оценки вероятности победы кандидатов в каждом шта-

те и, следовательно, большинства голосов в Коллегии выборщиков. Экспериментальная проверка выполнена в ходе предвыборных опросов в ходе президентской кампании 2008 года, и прогноз победы Б. Обамы оказался верным.

В работе [67] на основе данных с 12 000 уникальных опросов граждан Испании и разработанных байесовских моделей для объединения опросов подтверждено мнение, что анализ результатов текущих опросов населения может стать основой для определения времени выборов и сроков принятия мер по популяризации действий правительства.

В работе [68] авторы предлагают математическую модель оценки потенциальной границы неопределившихся в решении за кого голосовать избирателей штата на федеральных выборах в США, характеризующихся наличием двухпартийной системы и соответственно предполагающим выбор одного из двух. Предлагаемая модель позволяет оценивать разницу в голосовании даже при отсутствии каких-либо статистических данных, основываясь только на экспертных оценках чисел. На основе предлагаемой модели нахождение рассматриваемой оценки сводится к нахождению максимин функции разности двух билинейных функций с одним и тем же первым векторным аргументом, вторые векторные аргументы которого принадлежат многограннику связанных переменных (стратегий кандидатов), и этот минимум ищется на другом многограннике. При определенных естественных предположениях это ожидание математически описывается нелинейной функцией на многограннике из связанных переменных, которые являются стратегиями кандидатов. Хотя нахождение значения этого ожидания может представлять вычислительные трудности, структура этой функции такова, что нахождение наилучшей гарантированной нижней оценки значения этого ожидания для конкретного кандидата от основной партии сводится к решению вспомогательной задачи квадратичного программирования специального состава. Это позволяет формулировать и решать масштабные задачи планирования, связанные с мобилизацией избирателей, не участвующих в голосовании, поскольку квадратичное программирование является одним из наиболее мощных методов математического программирования, и также доступно хорошо разработанное программное обеспечение, реализующее эти методы.

В итоге можно отметить, что структурные модели опираются на опросные данные, поэтому точность их прогнозов зависит не только от используемых математических методов, но и от качества опросных данных. Прогнозы, использующие опросы, варьируются от более упрощенных взвешенных средних значений опроса (агрегирование опросов с учетом размера опроса) до некоторых очень сложных динамических линейных моделей.

### 4.3. Гибридные модели

В [69] предлагается «Модель канцлера» для прогноза результата выборов в Бундестаг Германии. «Модель канцлера» учитывает комплексное влияние на результат выборов долгосрочных, краткосрочных и среднесрочных факторов. Учитываются три параметра:

- популярность действующего канцлера;
- долгосрочный баланс сторонников партии в немецком электорате, которые официально в партии не состоят (так называемых партизан);
- расходы действующего правительства.

«Модель канцлера» прогнозирует долю голосов правящих партий, находящихся в

коалиции. Как правило, двух партий: социал-демократы и зеленые; или христианские демократы и свободные демократы. Коалиция для большего числа партий не рассматривается. Коэффициенты параметров прогноза (предикторов) оцениваются, начиная с выборов с 1949 г. и по настоящее время. Первое тестирование «Модель канцлера» прошла на выборах 2002 года, затем в 2005 и 2008 годах, и все прогнозы были точными. Несмотря на достигнутые положительные результаты, в среде социологов и политиков сохраняется недоверие к предлагаемой модели, так как она является «академической» и во многом повторяет аналогичные варианты для американских выборов. Однако созданные авторами прогнозные методы не копируют американские. «Модель канцлера» учитывает цикличность прихода в правительство одних и тех же партий, которые заключают между собой достаточно постоянные коалиции для формирования правительства, следовательно, учитывается влияние не одной партии, а коалиции. Также авторы оценивают и влияние краткосрочных факторов, например, удовлетворение общества работой действующего правительства. Прогноз моделей для выборов в США обычно учитывает этот фактор, опираясь на показатели экономической эффективности и популярность президента. В прогнозных немецких моделях в качестве краткосрочных факторов учитывают экономические предикторы: безработица, дефицит бюджета, а не популярность партии или канцлера. Авторы «Модели канцлера» в качестве определяющего фактора учитывают только популярность канцлера. Это подтвердили и предварительные тесты, выполненные авторами «Модели канцлера», показавшие отсутствие ощутимого влияния при включении в оценку экономических факторов, таких как экономический рост, безработица или инфляция. Популярность канцлера является главным и единственным предиктором. Подобная специфика проведения парламентских выборов характерна для Германии. Традиционно в прогнозных моделях имеется только один победитель. В «Модели канцлера» для победы правящей партии необходима коалиция. Авторы [69] считают, что в идеале прогнозная модель результата выборов в многопартийной системе должна прогнозировать доли голосов для каждой из сторон, но создать такую модель сложно и не очень рентабельно, поэтому авторы сделали акцент рассмотрения на правящих партиях, что повышает точность получаемых прогнозов. Если правящие партии набирают достаточно голосов на выборах, чтобы получить большинство мест в бундестаге, они выигрывают избрание и пребывание в должности. Если нет, то можно рассмотреть другие комбинации коалиций. Необходимо определить «достаточную» долю голосов для победы на выборах, так как пропорциональность голосов определяет места партии на выборах при условии, что партия имеет более пяти процентов национального голосования. Чем больше объединенный пул «потраченных впустую» голосов, тем ниже требование на долю голосования меньше 50% для обеспечения большинства мест, т.е. недостаточно «выиграть битву», так как это не дает гарантии «управления».

Авторы [70] предложили методику прогнозирования результатов выборов, сочетающую структурную модель с постоянным агрегированием результатов опросов. Основной структурной модели является байесовская иерархическая структура, обобщающая данные на трех уровнях: районном, региональном и национальном. Далее авторы используют байесовскую стратегию, чтобы добавить в структурную модель результаты последних опросов. Таким образом, появляется возможность учета влияния партий с небольшими историче-

скими данными, в том числе и для партий, участвующих в выборах впервые. Предлагаемая модель опробована на результатах выборов в Конгресс Испании 2015 года, когда две новые политические партии получили 30% голосов. И такая ситуация становится все более привычной для Европы после 2008 года. Научный подход к прогнозированию выборов основан на статистическом моделировании, использовании опросов о намерениях голосования или симпатиях выборщиков, использовании рынков политического прогнозирования на основе ставок для кандидатов. Новизна предлагаемой модели заключается в ее адаптации к ситуациям, когда недостаточно исторических данных для применения гибридных методов. В этом случае результаты выборов определяются местами, выигранными на местном уровне. В среднем по стране, это дает предсказание представительства партии в парламенте. Описанный подход заключается в том, чтобы получить средний показатель популярности по стране для каждой партии и определить, как этот национальный показатель переходит на региональный и городской уровни. Для того чтобы иметь дело с небольшими (или отсутствующими) историческими данными используется структурная модель намерения голосования, обученная на «глубоких» данных, полученных в предвыборных опросах. В Испании они выполняются финансируемым правительством исследовательским центром СНГ и позволяют оценить взаимосвязь между географическими и демографическими характеристиками и выбором избирателей. Недостатком этих наборов данных является то, что размер выборки в некоторых регионах очень низкий, и образец может быть не репрезентативным. Авторы [70] решают эту проблему, используя данные переписи населения. Описанная модель синтезирована с моделью опроса, эффективно вычисляющей средние значения опубликованных опросов и исправляющей потенциальные источники предвзятости, такие как различное качество методологий опроса, а также изменение временных промежутков опросов по мере приближения даты выборов. Из-за отсутствия длинных исторических данных синтез выполняется на основе байесовского подхода синтеза доказательств.

Таким образом, фундаментальная модель производит моделирование локальных результатов для каждой партии; затем полученные результаты преобразуются в количество возможных выборных мест на уровне региона или города (локальные результаты). Локальные результаты агрегируются на национальном уровне для каждого моделирования; получают вес, соответствующий тому, как близко подразумеваемый средний национальный показатель соответствует прогнозируемым моделям опросов. Строятся распределения мест в парламенте. Каждому предполагаемому общенациональному распределению мест присваивается вес и средневзвешенные значения рассчитываются для формирования прогнозов. Структурная модель и модель опроса являются многоуровневыми регрессионными моделями.

#### 4.4. Выводы

Электоральные процессы, играющие исключительно важную роль в динамике развития общественных систем с демократической формой правления, относятся к числу наиболее сложных общественных процессов. Особенностью электоральных процессов являются устойчивая пространственно-временная регулярность (выборы происходят через определенный промежуток времени на фиксированной территории, имеющей более или менее сложившуюся структуру избирательных округов) и одновременное участие в них больших социально различных групп населения с широким спектром социологических параметров.

Последнее приводит к тому, что объективным фактором, определяющим характер и результат электоральных процессов, является статистический, связанный с действием больших групп населения, а к субъективному фактору можно отнести сознательную деятельность участвующих в выборах людей. Статистический фактор поддается анализу и может быть использован для количественного прогноза, а субъективный вносит стохастичность и непредсказуемость в выполнение и результат электорального процесса. Поэтому на сегодняшний день нет общепризнанной методики анализа и прогнозирования электоральных процессов. Построено множество моделей, объясняющих развитие электоральных процессов «задним числом», после того, как выборы прошли, и результат известен. Прогноз предполагает положения, явно коррелирующие с теми событиями, которые будут наблюдаться на практике после его составления. В этом расхождение между целями прогнозирования «академиков» и «партийных аналитиков». Первые дают качественные прогнозы изменения тренда, вторым нужны количественные результаты. Наиболее успешными являются модели, сочетающие традиционные методы анализа временных рядов с методами, учитывающими особенности человеческого фактора – возможности самоорганизации и наличие памяти о предыдущих состояниях системы.

### **Заключение**

В представленном литературном обзоре описаны направления перспективных исследований в области анализа и моделирования временных рядов для описания процессов в сложных экономических, социотехнических и социальных системах.

На основе проведённого анализа рассмотрены достоинства и недостатки существующих моделей и методов.

Это позволило сделать вывод, что для описания процессов в сложных системах с наличием человеческого фактора необходимо разрабатывать и использовать методы и инструментальные средства учета возможности самоорганизации человеческих групп и наличия памяти о предыдущих состояниях системы.

### **Литература:**

1. Авзалова Э.И. Интернет-коммуникации в избирательной кампании США. *Известия Иркутского государственного университета. Серия «Политология. Религиоведение»*. 2017;22:185-194.
2. Курс Д. Парламентские выборы в Израиле 2015 г.: причины, результаты и тенденции. 8 июня 2015. Институт Ближнего Востока. [Электронный ресурс]: режим доступа – свободный (дата обращения: 14.11.2019). URL: <http://www.iimes.ru/?p=24736>
3. Родин А. Небесная сеть: почему новый проект Маска — вызов для России. 17 февраля 2018 г. РБК. Мнение. [Электронный ресурс]: режим доступа – свободный (дата обращения: 14.11.2019). URL: [https://www.rbc.ru/opinions/technology\\_and\\_media/17/02/2018/5a86a1559a79473fe79a5813](https://www.rbc.ru/opinions/technology_and_media/17/02/2018/5a86a1559a79473fe79a5813)
4. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление: пер. с англ. М.: Мир, 1974. 248 с.
5. Льюис К.Д. Методы прогнозирования экономических показателей: пер. с англ. М.: Финансы и статистика, 1986. 130 с.
6. Боровков А.А. Математическая статистика. М.: Физматлит, 2007. 704 с. ISBN 978-5-94052-141-X
7. Кендалл М., Стюарт А. Статистические выводы и связи: пер. с англ. М.: Наука, 1973. 900 с.
8. Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды: пер. с англ. М.: Наука, 1976. 736 с.
9. Уилкс С. Математическая статистика: пер. с англ. М.: Наука, 1967. 632 с.
10. Кобзарь А.И. Прикладная математическая статистика. М.: Физматлит, 2006. 816 с. ISBN 5-9221-0707-0
11. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. М.: Наука, 1972. 248 с.

12. Foster F.G., Stuart A. A distribution-free test in time series dated on the breaking of records. *JRSS.* 1954;B16(1):1-22. <https://doi.org/10.1111/j.2517-6161.1954.tb00143.x>
13. Орлов Ю.Н., Шагов Д.О. Индикативные статистики для нестационарных временных рядов. Препринты ИПМ им. М.В. Келдыша. 2011;053:20.
14. Гнеденко Б.В. Курс теории вероятностей. М.: Физматлит, 1961. 406 с.
15. Шугай Ю.С. Нейросетевые алгоритмы прогнозирования событий и поиска предвестников в многомерных временных рядах. *Искусственный интеллект.* 2004;2:211–215.
16. Ермаков С.М. Метод Монте-Карло и смежные вопросы. М.: Наука, 1975. 471 с.
17. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. М.: Наука, 1982. 296 с.
18. Кожевников А.С. Программное обеспечение для статистического моделирования и анализа случайных процессов со скачками, описывающих динамику цен акций предприятий авиационной отрасли. *Труды МАИ.* 2012;59:230-242.
19. Khadjeh Nassirtoussi A., Ying Wah T., Ngo Chek Ling D. A novel FOREX prediction methodology based on fundamental data. *Afr. J. Bus. Manag.* 2011;5(20):8322-8330. <https://doi.org/10.5897/AJBM11.798>
20. Anastasakis L., Mort N. Exchange rate forecasting using a combined parametric and non-parametric self-organising modelling approach. *Expert Sys. Appl.* 2009;36(10):12001-12011. <https://doi.org/10.1016/j.eswa.2009.03.057>
21. Vanstone B., Finnie G. Enhancing stockmarket trading performance with ANNs. *Expert Sys. Appl.* 2010;37(9):6602-6610. <https://doi.org/10.1016/j.eswa.2010.02.124>
22. Vanstone B., Finnie G. An empirical methodology for developing stockmarket trading systems using artificial neural networks. *Expert Sys. Appl.* 2009;36(3):6668-6680. <https://doi.org/10.1016/j.eswa.2008.08.019>
23. Sermpinis G., Laws J., Karathanasopoulos A., Dunis C. L. Forecasting and trading the EUR/USD exchange rate with gene expression and psi sigma neural networks. *Expert Sys. Appl.* August 2012;39(10):8865-8877. <https://doi.org/10.1016/j.eswa.2012.02.022>
24. Huang S.-C., Chuang P.-J., Wu C.-F., Lai H.-J. Chaos-based support vector regressions for exchange rate forecasting. *Expert Sys. Appl.* 2010;37(12):8590-8598. <https://doi.org/10.1016/j.eswa.2010.06.001>
25. Premanode B., Toumazou C. Improving prediction of exchange rates using differential EMD. *Expert Sys. Appl.* 2013;40(1):377-384. <https://doi.org/10.1016/j.eswa.2012.07.048>
26. Bahrepour M., Akbarzadeh T. M.-R., Yaghoobi M., Naghibi S. M.-B. An adaptive ordered fuzzy time series with application to FOREX. *Expert Sys. Appl.* 2011;38(1):475-485. <https://doi.org/10.1016/j.eswa.2010.06.087>
27. Mabu S., Hirasawa K., Obayashi M., Kuremoto T. Enhanced decision making mechanism of rule-based genetic network programming for creating stock trading signals. *Expert Sys. Appl.* 2013;40(16):6311-6320. <https://doi.org/10.1016/j.eswa.2013.05.037>
28. Орлов Ю.Н., Осминин К.П. Нестационарные временные ряды: методы прогнозирования с примерами анализа финансовых и сырьевых рынков. М.: ЛИБРОКОМ, 2011. 384 с. ISBN 978-5-397-071272-7
29. Орлов Ю.Н., Федоров С.Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера–Планка. *ТРУДЫ МФТИ.* 2016;8(2):126-133.
30. Fuentes M. Non-Linear Diffusion and Power Law Properties of Heterogeneous Systems: Application to Financial Time Series. *Entropy.* 2018;20(9):649. <https://doi.org/10.3390/e20090649>
31. Gunawardana A., Meek C., Xu P. A Model for Temporal Dependencies in Event Streams. Microsoft Research. 2011. 8 p. <https://www.microsoft.com/en-us/research/publication/a-model-for-temporal-dependencies-in-event-streams/>
32. Suraj Singh Chouhan, Ravi Khatri. Data Mining based Technique for Natural Event Pre-diction and Disaster Management. *International Journal of Computer Applications (IJCA).* 2016;139(14):34-39. <https://doi.org/10.5120/ijca2016909102>
33. Radinsky K., Horvitz E. Mining the Web to Predict Future Events. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. ACM. 2013. P. 255-264. <https://doi.org/10.1145/2433396.2433431>
34. Chatrath A., Miao H., Ramchander S., & Villupuram S. Currency jumps, cojumps and the role of macro news. *J. Int. Money Financ.* February 2014;40:42-62. <https://doi.org/10.1016/j.jimonfin.2013.08.018>
35. Huang C.-J., Liao J.-J., Yang D.-X., Chang T.-Y., & Luo Y.-C. Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Sys. Appl.* 2010;37(9):6409-6413. <https://doi.org/10.1016/j.eswa.2010.02.078>
36. Robertson C., Geva S., & Wolff R. What types of events provide the strongest evidence that the stock market is affected by company specific news? In: Proc. Fifth Australasian Data mining conference (AusDM2006). 2006;61:145-153.

37. Preethi P.G., Uma V., Ajit Kumar. Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction. *Procedia Computer Science*. 2015;48:84-89.  
<https://doi.org/10.1016/j.procs.2015.04.154>
38. Gerber M.S. Predicting crime using Twitter and kernel density estimation. *Decis. Support Syst.* 2014;61(1):115-125.  
<https://doi.org/10.1016/j.dss.2014.02.003>
39. Квейд Э. Анализ сложных систем. М.: Советское радио, 1969. 512 с.
40. Баскакова Ю.М., Дёмин А.А., Лашук Н.Е., Терентьева Н.Н. Американские практики электорального прогнозирования. Доклад по результатам исследования ВЦИОМ. М.: ОАО «Всероссийский центр изучения общественного мнения», 2016. 40 с. ISBN 978-5-9905970-5-1
41. Материалы круглого стола Общественной палаты РФ «Математика на службе избирателей», Пресс-служба Общественной палаты РФ. [Электронный ресурс]: режим доступа – свободный (дата обращения: 14.02.2019). URL: <https://www.oprf.ru/press/news/2018/newsitem/44277>
42. Толстова Ю.Н. Математическое моделирование социальных процессов и социология. *Социологические исследования*. 2018;9:104-112.  
<https://doi.org/10.31857/S013216250001965-4>
43. Толстова Ю.Н. Соотнесение теоретического и эмпирического знания при использовании математических методов в социологическом исследовании. *Социологические исследования*. 2018;12:39-48.  
<https://doi.org/10.31857/S013216250003164-3>
44. Walther D. Picking the winner(s): Forecasting elections in multiparty systems. *Elect. Stud.* December 2015;40:1-13.  
<https://doi.org/10.1016/j.electstud.2015.06.003>
45. Kryshchanovskii A.O. The Russian population's attitude toward the president's activities. *Russian Politics & Law*. 1995;33(6):61-68.  
<https://doi.org/10.2753/RUP1061-1940330661>
46. Крыштановский А. Методы анализа временных рядов. *Мониторинг общественного мнения: экономические и социальные перемены (4М)*. 2000;2(46):44-51.
47. Крыштановский А.О. «Кластеры на факторах» – об одном распространенном заблуждении. *Социология: Методология, методы, математические модели*. 2005;21:172-187.
48. Крыштановский А.О. Ограничения метода регрессионного анализа. *Социология: Методология, методы, математические модели*. 2000;12:96-112.
49. Bonica A.A., Rosenthal H.B., Rothman D.J.C. The political polarization of physicians in the United States: An analysis of campaign contributions to federal elections, 1991 through 2012. *JAMA Internal Medicine*. 2014;174(8):1308-1317.  
<https://doi.org/10.1001/jamainternmed.2014.2105>
50. Panagopoulos C. The dynamics of voter preferences in the 2010 congressional midterm elections. *Forum*. 2010;8(4):Article number 9.  
<https://doi.org/10.2202/1540-8884.1402>
51. Чуров В.Е., Арлазаров В.Л., Соловьев А.В. Итоги выборов. Анализ электоральных предпочтений. *Труды Института системного анализа Российской Академии Наук*. 2008;38:6-22.
52. Foucault M., Nadeau R. Forecasting the 2012 French Presidential Election. *PS: Political Science & Politics*. April 2012;45(2):218-222.  
<https://doi.org/10.1017/S1049096512000066>
53. Healy A., Lenz G.S. Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*. January 2014;58(1):31-47.  
<https://doi.org/10.1111/ajps.12053>
54. Gormley I.C., Murphy T.B. A grade of membership model for rank data. *Bayesian Analysis*. 2009;4(2):265-296.  
<https://doi.org/10.1214/09-BA410>
55. Докторов Б. Главная проблема Обамы – не республиканцы, но экономика (Аналитики избирательной кампании активно изучают прошлое для предсказания будущего). 11 Сентября 2011. [Электронный ресурс]: режим доступа – свободный (дата обращения: 14.02.2019). URL: <https://fom.ru/special/kto-stanet-prezidentom-ssha/10161>
56. Erikson R.S.A., Panagopoulos C.B., Wlezien C.D. Likely (and unlikely) voters and the assessment of campaign dynamics. *Public Opin. Quart.* 2004;68(4):588-601.  
<https://doi.org/10.1093/poq/nfh041>
57. Dassonneville R. Electoral volatility, political sophistication, trust and efficacy: A study on changes in voter preferences during the Belgian regional elections of 2009. *Acta Politica*. 2012;47(1):18-41.  
<https://doi.org/10.1057/ap.2011.19>
58. Dewan T.A., Shepsle K.A.B. Political economy models of elections. *Annu. Rev. Polit. Sci.* 2011;14(1):311-330.  
<https://doi.org/10.1146/annurev.polisci.12.042507.094704>
59. Mundim P.S. The press and the vote in the 2002 and 2006 Brazilian presidential campaigns [Imprensa e voto nas eleições presidenciais Brasileiras de 2002 e 2006]. *Revista de Sociologia e Política*. 2012;20(41):123-147.  
<https://doi.org/10.1590/S0104-44782012000100009>

60. Hillygus S.D. The evolution of election polling in the United States. *Public Opin. Quart.* 2011;75(5):962-981. <https://doi.org/10.1093/poq/nfr054>
61. Cassino D. How Today's Political Polling Works. *Harvard Business Review*. August, 2016. <https://hbr.org/2016/08/how-todays-political-polling-works>
62. Wright Fred A., Wright Alec A. How surprising was Trump's victory? Evaluations of the 2016 U.S. presidential election and a new poll aggregation model. *Electoral Studies*. August 2018;54:81-89. <https://doi.org/10.1016/j.electstud.2018.05.001>
63. Erikson R.S., Wlezien C. Forecasting the Presidential Vote with Leading Economic Indicators and the Polls. *PS: Political Science & Politics*. October 2016;49:669-672. <https://doi.org/10.1017/S1049096516001293>
64. Kennedy C., Blumenthal M., Clement S., Clinton J.D., Durand C., Franklin C., McGeeney K. An Evaluation of the 2016 Election Polls in the United States. *Public Opin. Quart.* March 2018;82(1):1-33. <https://doi.org/10.1093/poq/nfx047>
65. Fisher S.D., Ford R., Jennings W., Pickup M., Wlezien C. From polls to votes to seats: Forecasting the 2010 British general election. *Electoral Studies*. June 2011;30(2):250-257. <https://doi.org/10.1016/j.electstud.2010.09.005>
66. Linzer D.A. Dynamic Bayesian forecasting of presidential elections in the states. *J. Am. Stat. Assoc.* 2013;108(501):124-134. <https://doi.org/10.1080/01621459.2012.737735>
67. Walther D., Hellström J. The verdict in the polls: how government stability is affected by popular support. *West European Politics*. 2019;42(3):593-617. <https://doi.org/10.1080/01402382.2018.1490598>
68. Belenky A.S., King D.C. A mathematical model for estimating the potential margin of state undecided voters for a candidate in a US Federal election. *Math. Comput. Model.* March 2007;45(5-6):585-593. <https://doi.org/10.1016/j.mcm.2006.07.007>
69. Norporth H., Gschwend T. The chancellor model: Forecasting German elections. *Int. J. Forecasting*. January–March 2010;26(1):42-53. <https://doi.org/10.1016/j.ijforecast.2009.02.008>
70. Montalvo J.G., Papaspiliopoulos O., Stumpf-Fétizon T. Bayesian Forecasting of Electoral Outcomes with new Parties' Competition. Barcelona Graduate School of Economics. Working Papers 1065. February 4, 2019. [Электронный ресурс]. <https://ideas.repec.org/p/bge/wpaper/1065.html>

## References:

1. Avzalova E.I. Internet Communication in the US Election Campaign. *Izvestiya Irkutskogo gosudarstvennogo universiteta. Seriya «Politologiya. Religiovedenie» = The Bulletin of Irkutsk State University». Series «Political Science and Religion Studies»*. 2017;22:185-194 (in Russ.).
2. Course D. Israeli Parliamentary Elections 2015: Causes, Results and Trends. June 8, 2015. Institute of the Middle East. G. <http://www.iimes.ru/?p=24736> (in Russ.).
3. Rodin A. Sky network: why the new Mask project is a challenge for Russia. February 17, 2018 RBC. Opinion. [Electronic resource]. [https://www.rbc.ru/opinions/technology\\_and\\_media/17/02/2018/5a86a1559a79473fe79a5813](https://www.rbc.ru/opinions/technology_and_media/17/02/2018/5a86a1559a79473fe79a5813) (in Russ.).
4. Boks D., Dzhenkins G. *Analiz vremennykh ryadov. Prognoz i upravlenie* (Time Series Analysis. Forecast and management). Moscow: Mir; 1974. 248 p. (in Russ.).  
[Box D., Dzhenkins G. *Time Series Analysis: Forecasting and Control*. San Francisco: Hold-en-Day; 1970.]
5. L'yuis K.D. *Metody prognozirovaniya ekonomicheskikh pokazatelei* (Methods of forecasting economic indicators). Moscow: Finansy i statistika; 1986. 130 p. (in Russ.).
6. Borovkov A.A. *Matematicheskaya statistika* (Mathematical statistics). Moscow: Fizmatlit; 2007. 704 p. (in Russ.) ISBN 978-5-94052-141-X
7. Kendall M., Styuart A. *Statisticheskie vyvody i svyazi* (Statistical conclusions and relationships). Moscow: Nauka; 1973. 900 p. (in Russ.).
8. Kendall M., Styuart A. *Mnogomernyi statisticheskii analiz i vremennye ryady* (Multidimensional statistical analysis and time series). Moscow: Nauka; 1976. 736 p. (in Russ.).  
[Kendall M.G., Stuart A. *The advanced theory of statistics* (London)]
9. Uilks S. *Matematicheskaya statistika* (Mathematical statistics). Moscow: Nauka; 1967. 632 p. (in Russ.).
10. Kobzar' A.I. *Prikladnaya matematicheskaya statistika* (Applied Mathematical Statistics). Moscow: Fizmatlit; 2006. 816 p. (in Russ.). ISBN 5-9221-0707-0
11. Kagan A.M., Linnik Yu.V., Rao S.R. *Kharakterizatsionnye zadachi matematicheskoi statistiki* (Characterization problems of mathematical statistics). Moscow: Nauka; 1972. 248 p. (in Russ.).
12. Foster F.G., Stuart A. A distribution-free test in time series dated on the breaking of records. *JRSS.* 1954;B16(1):1-22. <https://doi.org/10.1111/j.2517-6161.1954.tb00143.x>

13. Orlov Yu.N., Shagov D.O. Indicative statistics for non-stationary time series. *Preprinty IPM im. M. V. Keldysha = Keldysh Institute preprints*. 2011;053:20 (in Russ.).
14. Gnedenko B.V. *Kurs teorii veroyatnostei* (Probability theory course). Moscow: Fizmatlit, 1961. 406 p. (in Russ.)
15. Shugai Yu.S. Neural network algorithms for predicting events and searching for precursors in multidimensional time series. *Iskusstvennyi intellekt = Artificial*. 2004;2:211-215 (in Russ.).
16. Ermakov S.M. *Metod Monte-Karlo i smezhnye voprosy* (Monte Carlo method and related issues). Moscow: Nauka; 1975. 471 p. (in Russ.).
17. Ermakov S.M., Mikhailov G.A. *Statisticheskoe modelirovanie* (Statistical Modeling). Moscow: Nauka; 1982. 296 p. (in Russ.).
18. Kozhevnikov A.S. The software for statistical modeling and analysis of random processes with jumps, which describe the stock price dynamics of aircraft industry. *Trudy MAI*. 2012;59:230-242 (in Russ.).
19. Khadjeh Nassirtoussi A., Ying Wah T., Ngo Chek Ling D. A novel FOREX prediction methodology based on fundamental data. *Afr. J. Bus. Manag.* 2011;5(20):8322-8330.  
<https://doi.org/10.5897/AJBM11.798>
20. Anastasakis L., Mort N. Exchange rate forecasting using a combined parametric and non-parametric self-organising modelling approach. *Expert Sys. Appl.* 2009;36(10):12001-12011.  
<https://doi.org/10.1016/j.eswa.2009.03.057>
21. Vanstone B., Finnie G. Enhancing stockmarket trading performance with ANNs. *Expert Sys. Appl.* 2010;37(9):6602-6610.  
<https://doi.org/10.1016/j.eswa.2010.02.124>
22. Vanstone B., Finnie G. An empirical methodology for developing stockmarket trading systems using artificial neural networks. *Expert Sys. Appl.* 2009;36(3):6668-6680.  
<https://doi.org/10.1016/j.eswa.2008.08.019>
23. Sermpinis G., Laws J., Karathanasopoulos A., Dunis C. L. Forecasting and trading the EUR/USD exchange rate with gene expression and psi sigma neural networks. *Expert Sys. Appl.* August 2012;39(10):8865-8877.  
<https://doi.org/10.1016/j.eswa.2012.02.022>
24. Huang S.-C., Chuang P.-J., Wu C.-F., Lai H.-J. Chaos-based support vector regressions for exchange rate forecasting. *Expert Sys. Appl.* 2010;37(12):8590-8598.  
<https://doi.org/10.1016/j.eswa.2010.06.001>
25. Premanode B., Toumazou C. Improving prediction of exchange rates using differential EMD. *Expert Sys. Appl.* 2013;40(1):377-384.  
<https://doi.org/10.1016/j.eswa.2012.07.048>
26. Bahrepour M., Akbarzadeh T. M.-R., Yaghoobi M., Naghibi S. M.-B. An adaptive ordered fuzzy time series with application to FOREX. *Expert Sys. Appl.* 2011;38(1):475-485.  
<https://doi.org/10.1016/j.eswa.2010.06.087>
27. Mabu S., Hirasawa K., Obayashi M., Kuremoto T. Enhanced decision making mechanism of rule-based genetic network programming for creating stock trading signals. *Expert Sys. Appl.* 2013;40(16):6311-6320.  
<https://doi.org/10.1016/j.eswa.2013.05.037>
28. Orlov Yu.N., Osminin K.P. *Nestatsionarnye vremennye ryady: metody prognozirovaniya s primerami analiza finansovykh i syr'evykh rynkov* (Unsteady time series: forecasting methods with examples of analysis of financial and commodity markets). Moscow: LIBROKOM, 2011. 384 p. (in Russ.). ISBN 978-5-397-071272-7
29. Orlov Yu.N., Fedorov S.L. Nonstationary time series trajectories generation on the basis of the Fokker–Planck equation. *TRUDY MFTI = Trudy MIPT*. 2016;8(2):126–133 (in Russ.).
30. Fuentes M. Non-Linear Diffusion and Power Law Properties of Heterogeneous Systems: Application to Financial Time Series. *Entropy*. 2018;20(9):649.  
<https://doi.org/10.3390/e20090649>
31. Gunawardana A., Meek C., Xu P. A Model for Temporal Dependencies in Event Streams. Microsoft Research. 2011. 8 p.  
<https://www.microsoft.com/en-us/research/publication/a-model-for-temporal-dependencies-in-event-streams/>
32. Suraj Singh Chouhan, Ravi Khatri. Data Mining based Technique for Natural Event Prediction and Disaster Management. *International Journal of Computer Applications (IJCA)*. 2016;139(14):34-39.  
<https://doi.org/10.5120/ijca2016909102>
33. Radinsky K., Horvitz E. Mining the Web to Predict Future Events. In: Proceedings of the 6th ACM International Conference on Web Search and Data Mining. ACM. 2013. P. 255-264.  
<https://doi.org/10.1145/2433396.2433431>
34. Chatrath A., Miao H., Ramchander S., & Villupuram S. Currency jumps, cojumps and the role of macro news. *J. Int. Money Financ.* February 2014;40:42-62.  
<https://doi.org/10.1016/j.jimonfin.2013.08.018>
35. Huang C.-J., Liao J.-J., Yang D.-X., Chang T.-Y., & Luo Y.-C. Realization of a news dissemination agent based on weighted association rules and text mining techniques. *Expert Sys. Appl.* 2010;37(9):6409-6413.  
<https://doi.org/10.1016/j.eswa.2010.02.078>

36. Robertson C., Geva S., & Wolff R. What types of events provide the strongest evidence that the stock market is affected by company specific news? In: Proc. Fifth Australasian Data mining conference (AusDM2006). 2006;61:145-153.
37. Preethi P.G., Uma V., Ajit Kumar. Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction. *Procedia Computer Science*. 2015;48:84-89.  
<https://doi.org/10.1016/j.procs.2015.04.154>
38. Gerber M.S. Predicting crime using Twitter and kernel density estimation. *Decis. Support Syst.* 2014;61(1):115-125.  
<https://doi.org/10.1016/j.dss.2014.02.003>
39. Kveid E. Analiz slozhnykh system (Analysis of complex systems). Moscow: Sovetskoe radio; 1969. 512 p. (in Russ.).
40. Baskakova Yu.M., Demin A.A., Lashuk N.E., Terent'eva N.N. *Amerikanskije praktiki elektoral'nogo prognozirovaniya. Doklad po rezul'tatam issledovaniya VTsIOM* (American practices of electoral forecasting. Report on the results of a VTsIOM study). Moscow: OAO «Vserossiiskii tsentr izucheniya obshchestvennogo mneniya»; 2016. 40 p. (in Russ.). ISBN 978-5-9905970-5-1
41. *Materialy kruglogo stola Obshchestvennoi palaty RF «Matematika na sluzhbe izbiratelei», Press-sluzhba Obshchestvennoi palaty RF* (Materials of the round table of the Public Chamber of the Russian Federation “Mathematics in the Service of Voters”, Press Service of the Public Chamber of the Russian Federation). [Electronic resource]:  
<https://www.oprf.ru/press/news/2018/newsitem/44277> (in Russ.).
42. Tolstova Yu.N. Mathematical Modeling of Social Processes and Sociology. *Sotsiologicheskie issledovaniya = Sociological Studies*. 2018;9:104-112 (in Russ.).  
<https://doi.org/10.31857/S013216250001965-4>
43. Tolstova Yu.N. Correlation of Theoretical and Empirical Knowledge when Using Mathematical Methods in Sociological Research. *Sotsiologicheskie issledovaniya = Sociological Studies*. 2018;12:39-48 (in Russ.).  
<https://doi.org/10.31857/S013216250003164-3>
44. Walther D. Picking the winner(s): Forecasting elections in multiparty systems. *Elect. Stud.* December 2015;40:1-13.  
<https://doi.org/10.1016/j.electstud.2015.06.003>
45. Kryshtanovskii A.O. The Russian population's attitude toward the president's activities. *Russian Politics & Law*. 1995;33(6):61-68.  
<https://doi.org/10.2753/RUP1061-1940330661>
46. Kryshtanovskii A. Methods of time series analysis. *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny = Public Opinion Monitoring: Economic and Social Change*. 2000;2(46):44-51 (in Russ.).
47. Kryshtanovskii A.O. “Clusters on factors” is about one common misconception. *Sotsiologiya: Metodologiya, metody, matematicheskie modeli (4M) = Sociology: Methodology, Methods, Mathematical Modeling (4M)*. 2005;21:172-187 (in Russ.).
48. Kryshtanovskii A.O. Limitations of the regression analysis method. *Sotsiologiya: Metodologiya, metody, matematicheskie modeli (4M) = Sociology: Methodology, Methods, Mathematical Modeling (4M)*. 2000;12:96-112 (in Russ.).
49. Bonica A.A., Rosenthal H.B., Rothman D.J.C. The political polarization of physicians in the United States: An analysis of campaign contributions to federal elections, 1991 through 2012. *JAMA Internal Medicine*. 2014;174(8):1308-1317.  
<https://doi.org/10.1001/jamainternmed.2014.2105>
50. Panagopoulos C. The dynamics of voter preferences in the 2010 congressional midterm elections. *Forum*. 2010;8(4):Article number 9.  
<https://doi.org/10.2202/1540-8884.1402>
51. Churov V.E., Arlazarov V.L., Solov'ev A.V. Election Results. Analysis of Electoral Preferences. *Trudy Instituta sistemnogo analiza rossiyskoy akademii nauk = Proceeding of the Institute for Systems Analysis of the Russian Academy of Science*. 2008;38:6-22 (in Russ.).
52. Foucault M., Nadeau R. Forecasting the 2012 French Presidential Election. *PS: Political Science & Politics*. April 2012;45(2):218-222.  
<https://doi.org/10.1017/S1049096512000066>
53. Healy A., Lenz G.S. Substituting the end for the whole: Why voters respond primarily to the election-year economy. *American Journal of Political Science*. January 2014;58(1):31-47.  
<https://doi.org/10.1111/ajps.12053>
54. Gormley I.C., Murphy T.B. A grade of membership model for rank data. *Bayesian Analysis*. 2009;4(2):265-296.  
<https://doi.org/10.1214/09-BA410>
55. Doktorov B. *Glavnaya problema Obamy – ne respublikancy, no ekonomika (Analitiki izbiratel'noj kampanii aktivno izuchayut proshloe dlya predskazaniya budushchego)* (The main problem of Obama is not in the Republicans, but in the economy (Campaign analysts are actively studying the past to predict the future)). 11 Sept. 2011. [Electronic Resource]:  
<https://fom.ru/special/kto-stanet-prezidentom-ssha/10161>
56. Erikson R.S.A., Panagopoulos C.B., Wlezien C.D. Likely (and unlikely) voters and the assessment of campaign dynamics. *Public Opin. Quart.* 2004;68(4):588-601.  
<https://doi.org/10.1093/poq/nfh041>
57. Dassonneville R. Electoral volatility, political sophistication, trust and efficacy: A study on changes in voter preferences during the Belgian regional elections of 2009. *Acta Politica*. 2012;47(1):18-41.  
<https://doi.org/10.1057/ap.2011.19>

58. Dewan T.A., Shepsle K.A.B. Political economy models of elections. *Annu. Rev. Polit. Sci.* 2011;14(1):311-330. <https://doi.org/10.1146/annurev.polisci.12.042507.094704>
59. Mundim P.S. The press and the vote in the 2002 and 2006 Brazilian presidential campaigns [Imprensa e voto nas eleições presidenciais Brasileiras de 2002 e 2006]. *Revista de Sociologia e Política.* 2012;20(41):123-147. <https://doi.org/10.1590/S0104-44782012000100009>
60. Hillygus S.D. The evolution of election polling in the United States. *Public Opin. Quart.* 2011;75(5):962-981. <https://doi.org/10.1093/poq/nfr054>
61. Cassino D. How Today's Political Polling Works. *Harvard Business Review.* August, 2016. [Electronic Resource]: <https://hbr.org/2016/08/how-todays-political-polling-works>
62. Wright Fred A., Wright Alec A. How surprising was Trump's victory? Evaluations of the 2016 U.S. presidential election and a new poll aggregation model. *Electoral Studies.* August 2018;54:81-89. <https://doi.org/10.1016/j.electstud.2018.05.001>
63. Erikson R.S., Wlezien C. Forecasting the Presidential Vote with Leading Economic Indicators and the Polls. *PS: Political Science & Politics.* October 2016;49:669-672. <https://doi.org/10.1017/S1049096516001293>
64. Kennedy C., Blumenthal M., Clement S., Clinton J.D., Durand C., Franklin C., McGeeney K. An Evaluation of the 2016 Election Polls in the United States. *Public Opin. Quart.* March 2018;82(1):1-33. <https://doi.org/10.1093/poq/nfx047>
65. Fisher S.D., Ford R., Jennings W., Pickup M., Wlezien C. From polls to votes to seats: Forecasting the 2010 British general election. *Electoral Studies.* June 2011;30(2):250-257. <https://doi.org/10.1016/j.electstud.2010.09.005>
66. Linzer D.A. Dynamic Bayesian forecasting of presidential elections in the states. *J. Am. Stat. Assoc.* 2013;108(501):124-134. <https://doi.org/10.1080/01621459.2012.737735>
67. Walther D., Hellström J. The verdict in the polls: how government stability is affected by popular support. *West European Politics.* 2019;42(3):593-617. <https://doi.org/10.1080/01402382.2018.1490598>
68. Belenky A.S., King D.C. A mathematical model for estimating the potential margin of state undecided voters for a candidate in a US Federal election. *Math. Comput. Model.* March 2007;45(5-6):585-593. <https://doi.org/10.1016/j.mcm.2006.07.007>
69. Norpoth H., Gschwend T. The chancellor model: Forecasting German elections. *Int. J. Forecasting.* January–March 2010;26(1):42-53. <https://doi.org/10.1016/j.ijforecast.2009.02.008>
70. Montalvo J.G., Papaspiliopoulos O., Stumpf-Fétizon T. Bayesian Forecasting of Electoral Outcomes with new Parties' Competition. Barcelona Graduate School of Economics. Working Papers 1065. February 4, 2019. <https://ideas.repec.org/p/bge/wpaper/1065.html>

**Об авторах:**

**Андреанова Елена Гельевна**, кандидат технических наук, доцент, доцент кафедры корпоративных информационных систем Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Scopus Author ID: 57200555430, <http://orcid.org/0000-0001-6418-6797>

**Головин Сергей Анатольевич**, доктор технических наук, профессор, заведующий кафедрой математического обеспечения и стандартизации информационных технологий Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78).

**Зыков Сергей Викторович**, доктор технических наук, доцент, профессор Департамента программной инженерии факультета компьютерных наук Национального исследовательского университета «Высшая школа экономики» (101000, Россия, г. Москва, ул. Мясницкая, д. 20). Scopus Author ID: 36146486900, <http://orcid.org/0000-0002-2115-5461>

**Лесько Сергей Александрович**, кандидат технических наук, доцент кафедры «Управление и моделирование систем» Института комплексной безопасности и специального приборостроения ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). Scopus Author ID: 57189664364, <https://orcid.org/0000-0002-6641-1609>

**Чукалина Екатерина Романовна**, магистрант кафедры корпоративных информационных систем Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78).

*About the authors:*

*Elena G. Andrianova*, Cand. Sci. (Engineering), Associated Professor, Associate Professor of the Department of ERP, IT Institute, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow 119454, Russia). Scopus Author ID: 57200555430, <http://orcid.org/0000-0001-6418-6797>

*Sergey A. Golovin*, Dr. Sci. (Engineering), Professor, Head of Department of Mathematical Support and Standardization of Information Technologies, IT Institute, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow 119454, Russia).

*Sergey V. Zыkov*, Dr. Sci. (Engineering), Associated Professor, Professor of School of Software Engineering of Faculty of Computer Science, Higher School of Economics, National Research University (20, Myasnitskaya ul., Moscow 101000, Russia). Scopus Author ID: 36146486900, <http://orcid.org/0000-0002-2115-5461>

*Sergey A. Lesko*, Cand. Sci. (Engineering), Associate Professor of the Department «Management and Modeling of Systems», Institute of Integrated Security and Special Instrumentation, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow 119454, Russia). Scopus Author ID: 57189664364, <https://orcid.org/0000-0002-6641-1609>

*Ekaterina R. Chukalina*, student, IT Institute, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow 119454, Russia).

*Поступила: 09.01.2020; Получена после доработки: 06.03.2020; Принята к опубликованию: 29.06.2020.*