Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА и ГОСУДАРСТВЕННОЙ СЛУЖБЫ при Президенте Российской Федерации»

ИНСТИТУТ ЭКОНОМИКИ, МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

НАПРАВЛЕНИЕ 38.03.01 ЭКОНОМИКА

Группа ОБ-7281-20 Кафедра микроэкономики

Допустить к защите

заведующий кафедрой микроэкономики

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Левин М.И.

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_ г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ОЦЕНКА КРИВОЙ ФИЛЛИПСА ДЛЯ РОССИЙСКОЙ ЭКОНОМИКИ

студент-бакалавр

Анфимов Александр Дмитриевич /\_\_\_\_\_\_\_\_\_\_\_\_/\_\_\_\_\_\_\_\_\_\_\_/

*(подпись) (дата)*

научный руководитель выпускной

квалификационной работы

к.э.н., Девятов Алексей Евгеньевич /\_\_\_\_\_\_\_\_\_\_\_\_/\_\_\_\_\_\_\_\_\_\_\_/

*(подпись) (дата)*

МОСКВА

2024 г.

**Оглавление**

[Введение 3](#_Toc201412940)

# Введение

Непараметрическая регрессия – это метод моделирования зависимости между переменными, который не предполагает явно заданной формы функции (линейной, полиномиальной и т.д.). Основное преимущество такого подхода заключается в гибкости, то есть адаптации к структуре без жёстких допущений. Именно поэтому данный метод является важным и актуальным в современных реалиях. В эпоху больших данных, когда наблюдается рост сложности данных (нелинейность, выбросы и мультимодальность) он позволяет анализировать сложные и неизвестные зависимости, которые практически невозможно оценить с помощью стандартных параметрических методов. Особенно широкое применение данный метод находит в машинном обучении и таких прикладных областях как экономика и медицина. Более подробно актуальные направления рассмотрены в соответствующей главе исследования.

В качестве основных подходов к реализации данного метода можно выделить как классические непараметрические методы (ядерные регрессии, локально полиномиальные оценки и др.) так и байесовские (регрессия гауссовских процессов, байесовские аддитивные регрессионные деревья). Более подробно существующие методы, а также их преимущества и недостатки, разбираются в обзорной главе данной работы.

Стоит отметить пробел в исследованиях, на котором фокусируется данная работа, а именно отсутствие работ, которые прямо анализируют информационную эффективность существующих подходов к оцениванию непараметрической регрессии. В большинстве исследований акцент идёт на предсказательной способности или уровня подгонки модели, а также качестве извлекаемых знаний из выборки. Однако стандартные используемые метрики, такие как и , не отражают полноту извлекаемых данных.

Таким образом, данное исследование направлено на то, чтобы заполнить этот пробел, предлагая другой, теоретико-информационный подход, к оценке эффективности построения непараметрической регрессии. Цель данной работы заключается в сравнении существующих классических непараметрических и байесовских методов с точки зрения количества извлекаемой информации из данных и построении сравнительной шкалы на основе нескольких метрик. Практическая значимость заключается в том, что в работе также описаны рекомендации и особенности выбора методов в зависимости от типа задач и данных.

Структура работы представлена следующим образом. Первая часть работы посвящена обзору актуальных направлений применения и исследования непараметрической регрессии. Во второй части работы приводится обзор и описание существующих подходов к оценке непараметрической регрессии. В третьей части описываются модификации выбранных для сравнения моделей, а также информационные критерии. В четвёртой части описаны данные. В пятой представлены результаты оценки и сравнения. В шестой части приводится заключение исследования.

# Основная часть

## 1. Обзор литературы

### 1.1. Современное состояние области исследования непараметрических методов

Непараметрическая регрессия позволяет моделировать зависимость без предположений о форме зависимости. На текущий момент существует множество различных методов оценки такой модели. Однако всё же можно выделить несколько крупных категорий, а также рассмотреть самые популярные на сегодняшний день подходы. В данной главе, будет приведён краткий список методов – более подробное описание будет проведено в главе 3 (ПОТОМ ИСПРАВИТЬ).

В первую очередь, очевидно, напрашиваются классические непараметрические методы. Они хорошо структурированы и описаны в работе Станислава А.[[1]](#footnote-1) и Matias S.[[2]](#footnote-2). В качестве самого популярного можно выделить семейство ядерных методов или оценка Надарая-Ватсона. В ней используется взвешенное среднее с ядром K и полосой пропускания (шириной окна) h. Каждое ядро имеет свои преимущества и недостатки, однако более подробно они будут описаны в отдельной главе. В качестве более общего класса можно выделить методы локально-полиномиальной регрессии, кроме оценки Надарая-Ватсона сюда входит ещё локально-линейная регрессия.

Другим известным классическим непараметрическим методом является сплайновая регрессия. В нём зависимость моделируется как кусочно-полиномиальная функция, гладко склеенная в о точках, называемых узлами. Существует множество сплайновых методов: полиномиальные сплайны, кубические сплайны, P-сплайны, B-сплайны, естественные сплайны и др..

Наконец, довольно простых, но также применяемым методом является метод k ближайших соседей. Как понятно из названия, в данной оценке для предсказания используются значения k ближайших соседей. Кроме того, есть различные модификации и улучшения метода, основным из которых является взвешенный по расстоянию KNN, который в свою очередь может быть модифицирован выбором различных функционалов для расстояния.

### 1.2. Актуальные работы по практическому применению непараметрической регрессии

Непараметрическая регрессия играет важную роль в современной статистике и машинном обучении. Она часто применяется в задачах, где требуется анализировать сложные, малоизученные и нелинейные данные, так как не предполагает явно заданной функциональной формы зависимости.

В последние годы наблюдается рост вычислительных мощностей и развитие больших данных. Особенно сильно это заметно в таких областях, как машинное обучение, искусственный интеллект, медицина, экономика и финансы, поэтому непараметрическая регрессия часто применяется для решения задач из этих областей, о чём и пойдёт речь в этой главе.

В качестве примера работы в области медицины можно выделить статью Альберта Канона, Гианлуки Байо и Иоанна Манолопполоу[[3]](#footnote-3). В ней авторы рассматривают современные подходы к оценке гетерогенных и индивидуальных эффектов лечения (CATE/ITE) с использованием методов непараметрической регрессии. Такой подход позволяет убрать или уменьшить необходимость в традиционных рандомизированных экспериментах, что особенно актуально в сферах, где они невозможны. В статье используется два основных непараметрических подхода: Bayesian Additive Regression Trees (BART) , а именно Bayesian Causal Forest (BCF) и мета-алгоритмы (S-, T-, X-, R-Learners). Эти методы позволили исследователям без жёстких предположений о функциональной зависимости оценить сложную связь между лечением и ковариатами. Авторы отмечают, что гибкий подход был критичным для минимизации систематических ошибок в оценках причинных эффектов и подчерчивают преимущество такого подхода перед параметрическими методами в контексте высокой размерности данных, несбалансированных групп и смешивающих факторов. Результативность методов демонстрировалась на двух датасете. Сначала исследователи провели анализ влияния школьных обеденных программ на здоровье детей, а потом показали высокую эффективность методов на синтетических данных. Однако авторы указывают на ряд важных ключевых допущений, соблюдение которых необходимо для эффективного применения предложенных методов, а именно отсутствие скрытых смешивателей и перекрытие групп.

В области экономики непараметрическая регрессия также является относительно популярным методов, в основном, потому что изучаемые зависимости неизвестны и довольно сложны. Хорошим примером этому может послужить работа Мингминг Джанга и др.[[4]](#footnote-4). В статье исследуется влияние инвестиций в возобновляемую энергетику на выбросы в Китае. Зависимость исследуется с помощью непараметрической аддитивной регрессионной модели STIRPAT, позволяющей учитывать как линейные, так и нелинейные взаимосвязи. Таким образом, исследователи могут учесть как известную информацию о функциональной зависимости через линейные коэффициенты (влияние таких стандартных факторов, как ВВП, уровень индустриализации и тд.), так и неизвестную изучаемую часть (в частности, влияние инвестиций) с помощью непараметрической регрессии. Проведённый анализ показал, что инвестиции оказываю сложное влияние на выбросы: сначала увеличивают их из-за инфраструктурных затрат и низкой эффективности, потом уменьшают благодаря замещению ископаемого топлива, однако снова начинают увеличивать выбросы при дальнейшем росте из-за экологических проблем инфраструктуры и компенсации углём. Также были обнаружены нелинейные зависимости для экономического роста (монотонный рост выбросов) и урбанизации (обратная U-кривая). В заключение авторы работы подчёркивают важность применения непараметрической регрессии в контексте исследования скрытых паттернов в экономике, которые сложно измерить параметрическими методами, а также предлагают практические рекомендации к повышению эффективности “зелёных” инвестиций в Китае.

Другим примером эффективного применения непараметрической регрессии в экономике может послужить работа Тингтинг Ченга и др.[[5]](#footnote-5). В ней авторы прогнозируют доходность акций с помощью непараметрической регрессии и других моделей. Для этого в исследовании была разработана многопериодная модель, которая учитывает локально стационарные предикторы (например, соотношении цены и доходности, дивидендную доходность) и оценивает регрессию с помощью ядерных методов сглаживания. По результатам исследования непараметрическая регрессия и её модификации (МА1 и МА2) показали качество лучшее по сравнению с линейными моделями, историческим средним уровнем и методами машинного обучения. Особенно сильно видно превосходство модели в долгосрочной перспективе (3–12 месяцев) и во времена кризисов и рыночной нестабильности (например, кризис 2009 года). Авторы указывают на то, что модель улавливает сложные паттерны, включая нелинейные связи между предикторами и доходностью, а также адаптируется к изменяющимся условиям через временную локализацию. Более того, инвестиционная стратегия построенная на модифицированной непараметрической регрессии МА2 показала высокую доходность (до 16% годовых) при низком риске. Таким образом, исследователи делают вывод, что непараметрическая регрессия является мощным и актуальным инструментом для инвестиционных стратегий и анализа рынка акций за счёт гибкости и устойчивости к структурным сдвигам в данных, так как финансовая область весьма волатильна.

Основным выводом из данных работ является факт того, что непараметрическая регрессия на сегодняшний день является важным и мощным инструментом во многих прикладных задачах, требующих гибкости и адаптивности модели.

### 1.3. Современные направления исследования непараметрической регрессии

Одним из ключевых направлений исследования являются проблемы построения непараметрической регрессии в высокоразмерных пространствах. Здесь можно выделить работу Елены Брадик и др.[[6]](#footnote-6), в которой авторы изучают современные методы в описанных выше условиях, фокусируясь на корневой n-согласованноости оценок линейных функционалов (например, регрессионного наклона или среднего эффекта). В статье авторы модифицируют подходы, использующие двойное (debiased) машинное обучение и разреженные аппроксимации. Таким образом, предложенный подход позволил им оценить параметры при отсутствии знаний о структуре данных. В качестве результатов исследования авторы выделяют установление условий разреженности и гладкости , при которых достижима скорость , и построение устойчивых к гетероскедастичности оценок через регуляризованные представители Рисса. В завершение, авторы подчёркивают преимущество подобного подхода по сравнению с тамими методами как debiased Lasso и series estimation, а именно высокая эффективность в условиях слабых предположениях о данных. Таким образом, данная работа обобщает непараметрические идеи и показывает, как с помощью машинного обучения можно улучшить качество непараметрической регрессии в высокоразмерных пространствах, расширяя применимость метода в экономике и других сферах.

Большинство современных методов построения непараметрической регрессии связаны с машинным обучением, поэтому множество исследований прямо или косвенно посвящены именно этой теме. Однако машинное обучение и нейросети применяются в работах по-разному.

Одним из ключевых направлений исследования являются проблемы построения непараметрической регрессии в высокоразмерных пространствах. Здесь можно выделить работу Елены Брадик и др.[[7]](#footnote-7), в которой авторы изучают современные методы в описанных выше условиях, фокусируясь на корневой n-согласованности оценок линейных функционалов (например, регрессионного наклона или среднего эффекта). В статье авторы модифицируют подходы, использующие двойное (debiased) машинное обучение и разреженные аппроксимации. Таким образом, предложенный подход позволил им оценить параметры при отсутствии знаний о структуре данных. В качестве результатов исследования авторы выделяют установление условий разреженности и гладкости , при которых достижима скорость , и построение устойчивых к гетероскедастичности оценок через регуляризованные представители Рисса. В завершение авторы подчёркивают преимущество подобного подхода по сравнению с такими методами как debiased Lasso и series estimation, а именно высокая эффективность в условиях слабых предположениях о данных. Таким образом, данная работа обобщает непараметрические идеи и показывает, как с помощью машинного обучения можно улучшить качество непараметрической регрессии в высокоразмерных пространствах, расширяя применимость метода в экономике и других сферах.

Другой подход применяется в работе Юпинг Сонга и др.[[8]](#footnote-8). В своей статье авторы фокусируются на интеграции современных методов глубокого машинного обучения с классической ядерной непараметрической регрессией. Качество моделей изучается в контексте прогнозирования высокочастотных финансовых данных, которые сложно предсказывать из-за нелинейности, высокой волатильности, долговременной памяти и частых скачков. Авторы подмечают, что обычные подходы плохо справляются с этой задачей: традиционные эконометрические модели (ARIMA, GARCH) ограничены при работе с подобными данными из-за предположений о линейности и стационарности; машинное обучение (SVM, нейросети) справляются с нелинейностью, однако не улавливают долговременную память; глубокие нейросети хорошо прогнозируют, но не интерпретируемы. В работе предлагается гибридная модель NR-LSTM, объединяющая интерпретируемость непараметрической регрессии со способностью глубокой нейросети улавливать долговременные эффекты и краткосрочные аномалии. В статье непараметрическая регрессия используется для моделирования тренда, в то время как глубокая нейросеть корректирует остатки, повышая общую точность прогноза. Полученная модель эмпирически проверялась на разнообразных финансовых данных (CSI 300, FTSE 100, S&P 500) и показала значительное улучшение по сравнению с другими подходами по метрике средней абсолютной ошибки (MAE), демонстрируя устойчивость и универсальность. Таким образом, данная работа подчеркивает перспективность гибридных методов, так как они позволяют использовать сильные стороны как традиционных статистических подходов, так и современных алгоритмов ИИ для решения задач в условиях больших данных и неопределённости.

Очень важной в данном контексте является статья Minshuo Chen и др.[[9]](#footnote-9). В ней авторы проводят теоретический анализ построения непараметрической регрессии на низкоразмерных многообразиях с помощью глубоких нейросетей с функцией активации ReLU. В точной постановке задачи исследуется восстановление функции , заданной на мерном римановом многообразии , вложенным в пространство , по зашумлённым данным , где – независимый субгауссовский шум. В работе авторы доказывают, что ReLU-сети могут эффективно аппроксимировать любую гладкую или гёльдерову функцию на многообразии с нужной погрешностью. Однако самое главное, что полученная скорость сходимости для такого метода построения непараметрической регрессии зависит от внутренней размерности d, то есть сложности самой зависимости, а не от размерности внешнего пространства D. В заключение, авторы продемонстрировали на примере задачи обработки картинок из датасета ImageNet, что классическими подходами практически невозможно получить высокую точность из-за проклятья размерности, в том время как предложенный метод вполне осуществим. Работа развивает идеи исследований Schmidt-Hieber (2019) и Nakada & Imaizumi (2020) об адаптации нейросетей к свойствам данных и демонстрирует преимущество подходов на основе глубоких нейросетей перед классическими подходами в задачах с высокоразмерными, но структурированными данными, где моделируемые зависимости гораздо проще пространства, в котором они находятся.

Несмотря на популярность нейросететевого подхода также остаются модели фокусирующиеся на модификации классических подходов. Особенно актуальной в контексте построения непараметрической регрессии является статья Taha Hussein Ali и др.[[10]](#footnote-10). В своей работе авторы представляют новый метод оценки параметра ширины окна (полосы пропускания) в непараметрической регрессии Надарая-Ватсона. В то время как классические подходы предлагают использовать фиксированнную или переменную (локально оптимальную) ширину окна, авторы предлагают улучшенный подход, основанный на универсальном пороговом уровне, вычисленном с помощью вейвлета Добеши. Для оценки шума используется медиана абсолютных отклонений (MAD), а порог определяется как , где – количество вейвлет коэффициентов. Предложенный подход авторы сравнивали с методами, которые основаны на геометрическом среднем, арифметическом среднем, размахе и медиане, на синтетических данных (размеры выборок – 32, 64 и 128) и реальных (извержения гейзера Old Faithful). По результатам оценки улучшенный подход показал наименьшую MSE по сравнению с классическими методами при всех объемах выборки и значениях начальной полосы пропускания на всех датасетах. Таким образом, данный метод является очень перспективным, особенно на маленьких выборках – он адаптивно настраивает ширину окна в регрессии Надарая-Ватсона, тем самым улучшая прогнозную способность модели. Более того, авторы подчеркивают, что использование других вейвлетов и пороговых стратегий (например, кросс-валидации) может ещё улучшить метод под конкретную задачу.

1. Anatolyev, Stanislav. (2009). Nonparametric regression (in Russian). Quantile. 37-52. [↑](#footnote-ref-1)
2. Salibian-Barrera, Matias. (2022). Robust nonparametric regression: review and practical considerations. 10.48550/arXiv.2211.08376. [↑](#footnote-ref-2)
3. Alberto Caron, Gianluca Baio, Ioanna Manolopoulou, Estimating Individual Treatment Effects using Non-Parametric Regression Models: a Review, Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 185, Issue 3, July 2022, Pages 1115–1149, https://doi.org/10.1111/rssa.12824 [↑](#footnote-ref-3)
4. Mingming Zhang, Zikun Yang, Liyun Liu, Dequn Zhou, Impact of renewable energy investment on carbon emissions in China - An empirical study using a nonparametric additive regression model, Science of The Total Environment, Volume 785, 2021, 147109, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2021.147109. [↑](#footnote-ref-4)
5. Cheng, Tingting and Gao, Jiti and Linton, Oliver B. and Yan, Yayi, Nonparametric Predictive Regression for Stock Return Prediction 1 (February 05, 2025). Available at SSRN: https://ssrn.com/abstract=5126279 or http://dx.doi.org/10.2139/ssrn.5126279 [↑](#footnote-ref-5)
6. Bradic, Jelena & Chernozhukov, Victor & Newey, Whitney & Zhu, Yinchu. (2019). Minimax Semiparametric Learning With Approximate Sparsity. 10.48550/arXiv.1912.12213. [↑](#footnote-ref-6)
7. Bradic, Jelena & Chernozhukov, Victor & Newey, Whitney & Zhu, Yinchu. (2019). Minimax Semiparametric Learning With Approximate Sparsity. 10.48550/arXiv.1912.12213. [↑](#footnote-ref-7)
8. Yuping Song, Chunchun Cai, Dexiang Ma, Chen Li, Modelling and forecasting high-frequency data with jumps based on a hybrid nonparametric regression and LSTM model, Expert Systems with Applications, Volume 237, Part A, 2024, 121527, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.121527. [↑](#footnote-ref-8)
9. Chen, Minshuo & Jiang, Haoming & Liao, Wenjing & Zhao, Tuo. (2022). Nonparametric regression on low-dimensional manifolds using deep ReLU networks: function approximation and statistical recovery. Information and Inference: A Journal of the IMA. 11. 10.1093/imaiai/iaac001. [↑](#footnote-ref-9)
10. Taha Hussein Ali, Heyam Abd Al-Majeed Hayawi & Delshad Shaker Ismael

    Botani (2021): Estimation of the bandwidth parameter in Nadaraya-Watson kernel non-parametric

    regression based on universal threshold level, Communications in Statistics - Simulation and

    Computation, DOI: 10.1080/03610918.2021.1884719 [↑](#footnote-ref-10)