

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И  
ИНФОРМАТИКИ**

**Кафедра математического моделирования и анализа данных**

Аннотация к дипломной работе

**ПРИМЕНЕНИЕ ДИНАМИЧЕСКИХ ФАКТОРНЫХ  
МОДЕЛЕЙ ДИСКРЕТНЫХ МНОГОМЕРНЫХ  
ДАННЫХ**

Артюкевич Арина Ивановна

Научный руководитель:

канд. физ.-мат. наук

доцент С.Н. Сталевская

Минск, 2022

## РЭФЕРАТ

Дыпломная праца, 64 старонкі, 29 малюнкаў, 9 табліц, 11 крыніц, 1 датак.

Ключавыя словы: ФАКТАРНАЯ МАДЭЛЬ, МАДЭЛЬ ПУАСОНА, МАДЭЛЬ НЕГБІНА, ФАКТАРНЫ АНАЛІЗ.

*Аб'ект даследавання* — прымяненне дынамічных фактарных мадэляў дыскрэтных шматмерных дынных у метадах аналізу статыстычных сукупнасцей і прыкладных задач.

*Мэта работы* — вывучэнне асаблівасцяў прымянення дынамічных фактарных мадэляў і магчымасці іх выкарыстання ў прыкладных задачах, а ў прыватнасці ў задачах фільтрацыі, прагназавання, аналізу фактарных структур у кантэксце дынамічных шматмерных дыскрэтных мадэляў.

*Метады даследавання* — метады тэорыі верагоднасці і матэматычнай статыстыкі, фактарны аналіз, эканаметрычнае мадэляванне.

*Вынік* — разгледжана Р-тэхніка фактарнага аналізу, пабудаваны мадэлі Пуасона і Нэгбіна для дыскрэтных шматмерных даных.

*Вобласць прымянення* — фактарны аналіз дыскрэтных шматмерных даных, зніжэнне памернасці, фільтраванне.

## ABSTRACT

Graduate work, 64 pages, 29 drawings, 9 tables, 11 sources, 1 attachment.

Key words: FACTOR MODEL, POISSON MODEL, NEGBIN MODEL, FACTOR ANALYSIS.

*Object of research* — application of dynamic factor models for multivariate count data in methods of analysis of statistical aggregates and applied problems.

*Purpose of the work* — to study the features of the application of dynamic factor models and the possibility of their use in applied problems, in particular in the problems of filtering, forecasting, analysis of factor structures in the context of dynamic multidimensional discrete models.

*Research methods* — methods of probability theory and mathematical statistics, factor analysis, econometric modeling.

*Result* — P-technique of factor analysis is considered, Poisson and Negbin models are constructed for discrete multidimensional data.

*Scope* — factorial analysis of discrete multidimensional data, dimensionality reduction, filtering.

## РЕФЕРАТ

Дипломная работа, 64 страниц, 29 рисунков, 9 таблиц, 11 источников, 1 приложение.

Ключевые слова: ФАКТОРНАЯ МОДЕЛЬ, МОДЕЛЬ ПУАССОНА, МОДЕЛЬ НЕГБИНА, ФАКТОРНЫЙ АНАЛИЗ.

*Объект исследования* — применение динамических факторных моделей дискретных многомерных данных в методах анализа статистических совокупностей и прикладных задач.

*Цель работы* — изучение особенностей применения динамических факторных моделей и возможности их использования в прикладных задачах, а в частности в задачах фильтрации, прогнозирования, анализа факторных структур в контексте динамических многомерных дискретных моделей.

*Методы исследования* — методы теории вероятности и математической статистики, факторный анализ, эконометрическое моделирование.

*Результат* — рассмотрена Р-техника факторного анализа, построены модели Пуассона и Негбина для дискретных многомерных данных.

*Область применения* — факторный анализ дискретных многомерных данных, снижение размерности, фильтрация.

## ВВЕДЕНИЕ

Будет рассматриваться динамическая факторная модель для анализа многомерных данных дискретного временного ряда, где дискретный временной ряд — это временной ряд с целочисленными компонентами. Такая модель учитывает характерные, а также общие последовательно коррелированные латентные факторы, чтобы учесть потенциально сложную динамическую взаимозависимость между дискретными временными рядами. Модель оценивается с использованием альтернативных распределений (Пуассона и отрицательного бинома). Оценка максимального правдоподобия (ML) требует численного интегрирования высокой размерности. Процедура интегрирования Монте-Карло (MC), известная как эффективная выборка по важности (EIS), которая дает быстрые и точные численные оценки функции правдоподобия. Модель применяется к данным временных рядов, состоящих из некоторого количества сделок с 5-минутными интервалами для пяти акций торгуемых на Нью-Йоркской фондовой бирже (NYSE) из двух промышленных секторов. Предполагаемая модель учитывает все ключевые особенности динамики и распределения данных. Наличие общего фактора интерпретируется как отражение новостей рынка.

Рассматриваемая модель является легкой для интерпритации факторной моделью для многомерных регрессионных рядов. Данная модель основывается на построенных ранее моделях Йоргенсена (1999) и Хельда (2003). Первая модель включает в себя всего один общий динамический фактор без уникальных компонент. Вторая модель является статической многомерной факторной моделью Пуассона. Исходная модель допускает общую серийную корреляцию, равно как и уникальные факторы условного среднего значения распределения. В силу этого, она может показывать нетривиальные одновременные и сезонные взаимодействия между числовыми рядами. Она также может учитывать различные предположения о распределении для условного распределения наблюдений с учетом факторов. Это может иметь решающее значение, поскольку обычно используемое распределение Пуассона имеет индекс рассеивания, равный единице (последний определяется как отношение между дисперсией и средним значением). Однако многомерные данные часто демонстрируют сильное избыточное рассеивание (индекс значительно больше единицы), который не может быть полностью учтенным условным распределением Пуассона. Следовательно, важно учитывать условные распределения, которые могут содержать избыточное рассеивание, такие как отрицательный

бином и распределение двойного Пуассона.

Модель нелинейно зависит от ее динамических скрытых факторов. Следовательно, оценка правдоподобия требует численного интегрирования с высокой размерностью, для чего будет использоваться процедура EIS, разработанная Ричардом и Чжаном(2007). EIS — это универсальная, гибкая и простая в реализации процедура интегрирования МС, специально разработанная для обеспечения максимальной числовой точности. Это также облегчает изучение альтернативных спецификаций модели, которые обычно требуют лишь незначительных изменений базовой реализации EIS. И последнее, но не менее важное: EIS можно использовать для вычисления отфильтрованных и/или сглаженных оценок самих скрытых факторов. Некоторые статистические данные диагностических тестов основаны на таких оценках.

Затем построенная модель будет применяться к многомерному временно-му ряду, состоящему из сделок с 5-минутными интервалами для пяти акций, торгуемых на NYSE. В этом контексте на сделки напрямую влияет поступление новой информации, будь то конкретная акция (уникальный фактор), отрасль (фактор сектора) или рынок (рыночный фактор).

Также будет рассматриваться применение факторного анализа с целью выявления скрытых факторов с помощью Р-техники.

Выделение общих факторов и сжатие информации в ходе факторного анализа сводится к воспроизведению с той или иной степенью точности исходной корреляционной матрицы, т.е. предполагается, что редуцированная корреляционная матрица получена с использованием тех же объектов, но описанных меньшим числом переменных. Р-техника, предполагающая факторный анализ результатов экспериментальных исследований, выполненных на одном и том же объекте в различные промежутки времени.

Будет использоваться имитационный набор данных с характеристиками двух человек.