

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
Севастопольский национальный технический университет

В.В. ХОХЛОВ

**МНОГОМЕРНЫЙ ФАКТОРНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ
БАНКОВСКИХ ДЕПОЗИТОВ**

Севастополь 2009

ББК 33.24:В17
Х89
УДК 336:519.2

Рецензенты:

д-р экон. наук, проф., заведующий кафедрой Днепропетровской государственной финансовой академии А.Г. Бабенко;
д-р техн. наук, проф., заведующий кафедрой Севастопольского национального технического университета А.В. Цуканов;
канд. экон. наук, доцент, ректор Крымского института экономики и хозяйственного права С.В. Тумакова.

Научный редактор – канд. экон. наук, доцент, заместитель декана факультета Севастопольского национального технического университета Т.М. Одинцова.

Хохлов В.В.

Х89 Багатомірний факторний аналіз часових рядів банківських депозитів: монографія /В.В. Хохлов. – Севастополь: Вид-во СевНТУ, 2009. – 204 с.
ISBN 978-966-2960-42-6.

У монографії розглянуті й вирішені за допомогою нових підходів основні проблеми багатомірного факторного аналізу. Запропонована авторегресійно-факторна модель багатомірного часового ряду, за допомогою якої зроблений максимально правдоподібний прогноз значень залишків на депозитних банківських рахунках.

Хохлов В.В.

Х89 Многомерный факторный анализ временных рядов банковских депозитов: монография /В.В.Хохлов. – Севастополь: Изд-во СевНТУ, 2009. – 204 с.
ISBN 978-966-2960-42-6.

В монографии рассмотрены и решены с помощью новых подходов основные проблемы многомерного факторного анализа. Предложена авторегрессионно-факторная модель многомерного временного ряда, с помощью которой сделан максимально правдоподобный прогноз значений остатков на депозитных банковских счетах.

Рекомендовано к печати Ученым Советом СевНТУ, протокол № 7 от 05.03.2009 г.

ISBN 978-966-2960-42-6

ББК 33.24:В17

© Издательство СевНТУ, 2009

СОДЕРЖАНИЕ

Предисловие	4
1. Современный факторный анализ.....	9
1.1. Место факторного анализа среди эконометрических методов	9
1.2. Модель факторного анализа	29
1.3. Методы оценок параметров факторной модели	51
1.4 Проблемы факторного анализа.....	66
2. Новый подход к решению проблем факторного анализа.....	77
2.1. Оценка общности	77
2.2. Оценка факторных нагрузок ортогональной факторной модели.....	90
2.3. Оценка значений факторов	103
2.4. Исследование формирования остатков на депозитных счетах .	121
3. Факторный анализ многомерных временных рядов.....	131
3.1. Классический анализ временных рядов	131
3.2. Основные методы в анализе многомерных временных рядов..	149
3.3. Факторный подход к модели многомерных временных рядов .	168
3.4. Факторная модель в анализе многомерных временных рядов .	178
3.5. Факторный анализ в прогнозировании многомерных временных рядов.....	189
Библиографический список	200

«Все реки текут в море, но море не переполняется: к тому месту, откуда реки текут, они возвращаются, чтобы опять течь».

Екклесиаст 1, 7.

Предисловие

Познание законов движения материального мира – бесконечно. Невозможно даже представить, что в какой-то момент можно будет сказать: «Наконец-то мы знаем все и больше тайн мироздания для нас не существует». Очевидно, что такое допущение лишено здравого смысла. Пока в этом мире существует человечество, оно постоянно будет заглядывать под покров неведомого. При этом само движение народов по пути всеобщего материального изобилия создает новые и новые нерешенные проблемы. И на этом бесконечном пути должны и будут появляться новые методы и инструменты познания.

Поведение современной глобальной экономики, и, особенно, глобальной финансовой системы, становится все менее предсказуемой. Дефицит ресурсов в одной части мира, неадекватное поведение рыночных игроков в финансовом сегменте, перепроизводство товаров в Поднебесной и несчетная масса прочих причин порождают процессы, которые, в конечном итоге, сказываются без исключения на всех, живущих на нашей планете. Существующие эконометрические методы предсказания не смогли спрогнозировать момент начала и характер протекания глобального экономического кризиса. Создается впечатление, что экономико-математические методы анализа и исследования существующей реальности, призванные вовремя просигнализировать о надвигающихся проблемах, не могут разглядеть нечто важное, несомненно значительное, заставляющее срабатывать некие механизмы, работа которых может привести к экономическому хаосу. Математика, безусловно, не может

измерить Проведение Господне, но она должна стать инструментом фиксации скрытых от непосредственного наблюдения процессов, определяющих развитие и характер поведения экономических систем.

Основная масса методов проведения эконометрических исследований не обладает таким свойством – распознать незримое, эти методы в лучшем случае могут помочь выявить очевидные закономерности, так как на предварительном этапе проведения исследования должна быть постулирована модель исследуемого объекта. А поскольку конструирование модели возможно лишь на основе предварительных предположений о свойствах объекта, и если имеются лишь массивы необработанной статистической информации, то исследование на самой начальной стадии изучения нового явления может зайти в тупик. Поэтому должны появиться методы, которые позволят не только обработать статистические данные с целью выявления полезной информации, но и помогут сконструировать модель, адекватно описывающую реальную экономическую систему.

В настоящее время такие гуманитарные науки, как социология и психология достаточно широко используют многомерный факторный анализ обработки статистической информации, позволяющий выявить непосредственно неизмеряемые факторы. Этот метод как раз и дает возможность до некоторой степени раскрыть механизмы взаимодействия различных частей исследуемого объекта. Причем сам объект в социологических и психологических исследованиях – не техническая система, поведение которой можно описать аналитически, он – сверхсложная стохастическая система, поведение которой зачастую не объясняется видимыми причинами. И факторный анализ – фактически, единственный метод, позволяющий результативно исследовать такие объекты.

В эконометрических исследованиях факторный анализ практически не используется, и тому имеется несколько причин. Самая главная причина – идентификация решения, которая при традиционных методах факторного анализа допускает неоднозначность: нагрузки факторов при переменных могут

быть с помощью методов «вращения» преобразованы к другому виду, а ведь так – факторных решений может быть бесчисленное множество. Другая причина – плохие вычислительные свойства методов оценивания параметров модели, причем методов, которые должны давать, казалось бы, самые надежные результаты. Так, итерационная процедура получения максимально правдоподобных оценок по традиционной методике сходится в исключительно редких случаях. Поэтому на практике используют очень простые методы поиска параметров факторной модели, которые применялись еще в середине прошлого века без привлечения вычислительной техники. В какой степени можно доверять решениям, полученным этими методами, – вопрос открытый, но в эконометрическом исследовании, целью которого является улучшение материальной стороны жизни людей, даже предположение о неточности решения является недопустимым.

Нерешенные или решенные не до конца некоторые проблемы факторного анализа составляют еще одну группу причин, по которым этот метод математической статистики не задействован в исследованиях экономической сферы в должной степени. К таким проблемам можно отнести проблему определения числа факторов, проблему общности исследуемых показателей, проблему вычисления значений факторов и проблему получения однозначного факторного решения. Преодолению некоторых «белых пятен» многомерного факторного анализа посвящена данная работа.

В ней предпринята попытка привлечь внимание экономистов и финансистов к возможностям факторного анализа в исследовании причинных процессов для нетривиального понимания экономических явлений. Именно многомерный факторный анализ позволяет выявить скрытые механизмы формирования долгосрочных тенденций в глобальной экономике, поэтому использование его методов очень перспективно для исследования экономических систем, причем как на микро, так и на макроуровне.

В первом разделе рассмотрены особенности факторной модели, имеющей как общие, так и характерные факторы, дается систематическое

описание последних достижений в факторном анализе, а также разобраны проблемы факторного анализа выборок с независимыми наблюдениями.

Второй раздел посвящен решению проблем многомерного факторного анализа. В нем предложен новый подход к получению оценок общностей переменных, описывающих экономический объект, приводится принципиально новый метод получения максимально правдоподобных оценок параметров факторной модели, отвечающих принципу простой структуры и дающих однозначное решение. Во втором разделе решена проблема факторов, т.е. предложен метод получения оценок значений факторов, соответствующих основным постулатам факторного анализа. С помощью разработанных методов сконструирована факторная модель, на основании которой рассматриваются особенности формирования остатков на депозитных счетах коммерческого банка в зависимости от уровня процентных ставок и их изменения.

Факторный анализ представляет собой не только самостоятельный способ познания материального мира, но он может выступать в роли инструмента, позволяющего решить некоторые проблемы в смежных разделах математической статистики. А именно в решении проблем многомерных временных рядов.

Экономическая и финансовая информация представляет собой значительные объемы одновременных наблюдений за большим числом показателей (переменных). Величины этих показателей определяются двумя главными воздействиями. Первое определяется влиянием значений показателей в предшествующие моменты времени – так называемой авторегрессионной зависимостью. Второе воздействие определяется влиянием других переменных одно на другое – корреляционной зависимостью. К настоящему моменту не существует моделей многомерных временных рядов, которые учитывали бы обе группы воздействий. Создание такой модели позволило бы в значительной мере повысить точность выводов эконометрического исследования. Этому посвящен третий раздел работы.

В нем систематически изложены основные теоретические моменты анализа временных рядов как одномерных, так и многомерных, приведены последние достижения в этой области. И в этом разделе предложена принципиально новая модель многомерных временных рядов, учитывающая как авторегрессионные воздействия, так и корреляции между переменными. Она состоит из двух уравнений: первое описывает временные переходы значений факторов, непосредственно не наблюдаемых; второе представляет собой, по сути, факторную модель зависимости наблюдаемых переменных от скрытых факторов. В третьем разделе выведены оценки параметров такой модели, и, в частности, получена оценка матрицы переходов значений факторов. Сконструированная модель позволила предсказать значения остатков на депозитных счетах двух видов для национальной и иностранной валют. Предсказаны также значения процентных ставок, причем такой прогноз можно рассматривать, как рекомендации коммерческому банку в своей процентной политике.

Автор выражает признательность и благодарность А.П. Вожжову – без его внимания и рекомендаций написание этой работы было бы невозможно, а также А.В. Цуканову – дискуссии с ним навели автора на ряд свежих идей.

1. Современный факторный анализ

1.1. Место факторного анализа среди эконометрических методов

Глубокое исследование экономических процессов, явлений и объектов невозможно без привлечения методов и средств математической статистики, использование которых в экономической сфере объединяется под общим названием – «эконометрия».

Эконометрия является математическим аппаратом экономического анализа, с одной стороны. С другой стороны, в силу обширного числа различных методов ею используемых и развиваемых, эконометрия давно уже рассматривается как самостоятельная наука. И эта наука позволяет:

- всесторонне анализировать и оценивать сконструированные с помощью экономической теории и экономических подходов математические модели объекта исследования;
- на основе аппарата теории статистического вывода проверять различные гипотезы об объекте исследования, тем самым устанавливать истинность экономической теории;
- получать экономически обоснованные и статистически достоверные прогнозы развития экономических показателей, что является научной базой принятия решений на различных уровнях: начиная от обоснования хозяйственного решения на предприятии, и заканчивая выбором стратегии экономического развития государства.

Эконометрия является сочетанием двух областей знаний: экономической теории, которая часто использует математические формулировки, и математической статистики. Эконометрия позволяет получить числовое решение практической проблемы, причем в ее основе лежат не любые теоретические допущения, а подлинные факты – статистика наблюдений за реальным объектом. Именно ориентация на достоверность происходящих событий является главным стержнем эконометрического подхода к

исследованию экономического явления. И полное соответствие модели, получаемой как итог изучения экономического объекта, к действительности представляет собой главный критерий качества эконометрического исследования.

К особенностям использования эконометрии следует отнести:

- 1) большое число экономических показателей, описывающих процесс или объект и являющихся случайными величинами;
- 2) взаимосвязь или коррелированность этих переменных между собой;
- 3) неизвестная природа формирования случайности во взаимосвязях между показателями.

Первая особенность определяется системным подходом, характерным для эконометрии. Согласно этому подходу любой процесс или объект должен быть отображен в описании во всем многообразии как внутренних, так и внешних связей. Безусловно, полная детализация структуры явления приведет ее к бесконечности. И все же, на первом этапе исследования ни один аспект в функционировании явления, ни один показатель, характеризующий его особенности, не должен быть проигнорирован. Такой подход ведет к формированию огромного объема информации, обработка которого не мыслима без использования современных эконометрических методов.

Вторая особенность возникает в силу известного постулата, согласно которому все экономические переменные связаны между собой. Причем связь между двумя переменными может быть прямой: скажем, предложение некоторого продукта напрямую влияет на его цену. Но связь между переменными может быть и опосредственной: затраты на рекламу вряд ли непосредственно влияют на предложение продукта, но стимуляция спроса на него, благодаря рекламе, с неизбежностью приведет к росту предложения на этот товар.

Третья особенность определяется случайностью всего происходящего в этом мире, причем о природе происхождения случайности заранее ничего не известно. А именно, трудно наперед сказать, что определило числовое значение

экономического показателя в данный момент: то ли его значение в предыдущие моменты времени, то ли влияние на него в данный момент других факторов, а может быть и то и другое; а может быть влияние обстоятельства, не попавшего в поле зрения исследователя. Природа случайности несчетна, трудно даже предусмотреть все аспекты ее проявления.

Эконометрическое исследование традиционно имеет несколько этапов. На первом этапе, исходя из общеэкономических подходов, выделяются все возможные связи между экономическими переменными, а затем на основании статистического материала и, исходя из целей исследования, выбираются наиболее важные из них для дальнейшего включения в модель. Например, рыночный спрос и предложение должны объясняться соотношениями между объемом выпуска, объемом продаж, ценой предлагаемой, ценой запрашиваемой, потребительским доходом.

Включить в модель все переменные, чтобы объяснить поведение каждой из них, невозможно, поскольку всегда останется, хотя бы одна, внешняя переменная, не подлежащая объяснению, и это делает модель неполной или условной. Причем увеличением числа переменных и числа уравнений преодоление условности модели представляется невыполнимым.

На втором этапе выбранные соотношения группируются в модель, математические особенности которой зависят от характера экономических процессов. К примеру, модель спроса и предложения представляет собой систему одновременных уравнений: уравнения спроса, уравнения предложения и уравнения реакции рынка. Это линейные уравнения первой степени, но со стохастическими составляющими. Модели исследования финансового риска предполагают использование аппарата теории временных рядов, и, в частности, рядов Фурье. Динамические модели трудно представить без применения дифференциальных или разностных уравнений.

Неуклонное следование требованиям эконометрического подхода приводит к тому, что все эконометрические модели обладают общими свойствами:

- поведение экономически переменных определяется с помощью совместных и одновременных операций с некоторым конечным числом математических соотношений;
- модель, допускающая упрощение действительности, тем не менее, улавливает главные характерные черты изучаемого объекта;
- на основе достигнутого с ее помощью понимания реальной системы всегда можно предсказать ее будущее движение и управлять этим движением.

На третьем этапе статистические данные используются для оценки параметров модели. Причем выбор метода получения оценок также является достаточно ответственным моментом эконометрического исследования. Этот метод должен дать наилучшие для данного фактического материала оценки. Традиционно, к таким статистическим методам относятся метод максимального правдоподобия и метод наименьших квадратов. Хотя специфика статистических данных может сделать невозможным применение одного из них. К примеру, при наличии мультиколлинеарности в исходных данных использование метода наименьших квадратов для получения оценок регрессии приведет к существенным искажениям картины реального объекта и к фатальным ошибкам в принятии решений на основе такой модели.

На последнем этапе осуществляется проверка оцененной модели на адекватность к действительности. На основании ретроспективного прогноза делается вывод о достаточной реалистичности получаемой с ее помощью картины объекта и возможности использования ее в достижении целей исследования, или модель признается неадекватной. В последнем случае эконометрическое исследование должно быть воспроизведено с более ранних этапов, чтобы в конечном итоге получить соответствие модели и реального объекта.

К эконометрическим относится обширный круг методов математической статистики. Первым из них был классический регрессионный анализ. Применение этого метода обширно, но в экономике оно ограничивается достаточно жесткими предположениями о свойствах неизвестных случайных

возмущений со стороны внешних факторов, не объясняемых с помощью модели. Так, наличие гетероскедастичности (неоднородности дисперсии ошибок уравнения регрессии) или наличие автокоррелированности ошибок, или наличие сезонности в исходных данных, делает невозможным применение этого метода в исследовании экономических явлений.

Развитие регрессионного анализа как эконометрического метода, идет по пути преодоления противоречий между природой экономического объекта и допустимыми предпосылками использования математики в объяснении естества явлений.

Близко по идеологии к регрессионному анализу расположен другой важный эконометрический метод моделирования экономических процессов – использование систем одновременных уравнений. Однако в отличие от регрессионных моделей, в такой системе должна быть задана жесткая структура уравнений, отвечающая критериям идентифицируемости, причем одна и та же экономическая переменная может выступать как в роли результирующего показателя, так и в роли объясняющей величины, – в регрессионных уравнениях это не возможно. Формирование системы одновременных уравнений может быть проделано лишь на основании четких экономических представлений о природе изучаемого явления.

Регрессионный анализ и системы одновременных уравнений являются центральными методами эконометрии, однако, не менее важными являются другие разделы эконометрии: теория временных рядов (хотя этот раздел в силу своей важности и объемности рассматривается как самостоятельная наука), дискриминантный анализ, дисперсионный анализ; линейное, стохастическое, нелинейное и дискретное программирование, теория массового обслуживания, теория стохастических процессов, метод Монте-Карло, кластерный анализ и наконец, факторный анализ.

В отечественной экономической литературе очень часто под факторным анализом понимается раздел экономического анализа конкретных показателей, чаще всего на основании индексной мультипликативной модели. И в этой связи

такой факторный анализ не принято считать отдельной научной дисциплиной. В математической статистике под факторным анализом принято понимать нечто иное. А именно, словарь-справочник «Математика и кибернетика в экономике» [1] приводит такую формулировку: «Факторный анализ – направление многомерного анализа, которое исследует структуру матриц ковариаций и корреляций». Поэтому факторный анализ является ветвью математической статистики, и когда с его помощью исследуется экономический объект, факторный анализ становится эконометрическим методом. Однако полные возможности факторного анализа, как метода познания мира, в должной мере пока еще не реализованы. Попытаемся указать на его потенциал.

Наука является процессом, состоящим из непрерывного цикла построения моделей наблюдаемой действительности, которая вначале кажется необозримой в своем многообразии и не может быть хоть как-нибудь охвачена моделью. В естественных науках наряду с накапливанием наблюдений происходит формирование теоретических положений, которые затем проверяются на новых данных и постепенно совершенствуются.

Математика предоставляет возможность построения неограниченного числа моделей, с помощью которых можно точно описывать действительность. На ранних этапах научного познания реального мира приходилось полагаться на интуицию и гениальность отдельного исследователя, который составлял такую модель и проверял ее соответствие действительным условиям. Однако эти обе функции исследователя, а именно построение модели и проверку ее адекватности, можно систематизировать и формализовать.

Теоретические положения физики, химии и других естественных наук в значительной степени основаны на математических моделях. Для этих наук характерно то, что в условиях одного эксперимента в принципе можно проверить определенную закономерность путем манипулирования величинами в широких пределах, пользуясь их причинной обусловленностью. В науках, занимающихся изучением поведения человека, к которым относится

психология, социология, частично медицина и многие смежные области других отраслей знаний, где отсутствует четкая функциональная связь между причиной и следствием, один опыт не играет такой решающей роли. Присущие им закономерности можно обнаружить лишь путем статистической проверки результатов многих опытов, принимая или отклоняя определенные гипотезы. Процедура статистической проверки начинается с формулировки нулевой гипотезы («нет никакого различия») и альтернативной («имеет место различие»). Затем на основании выборки производится проверка нулевой гипотезы относительно альтернативной с помощью соответствующего критерия. Результатом статистической проверки является вывод о том, в скольких случаях на каждые 100 проведенных испытаний в предположении определенной модели отклонения от нее можно считать случайными. Может быть определена вероятность ошибочного принятия альтернативной гипотезы. Вероятность допустить ошибку при принятии нулевой гипотезы тоже имеет вполне определенное значение, называемое уровнем значимости. Таким образом, исследователь на заданном уровне значимости может остановиться на одной из этих гипотез. Имеется большой набор критериев значимости, с помощью которых осуществляется проверка различных гипотез. Выполнив процедуру проверки статистических гипотез, исследователь приходит к выводу о возможности принятия определенной модели. Для этой процедуры важно, на основании каких соображений исследователь определяет вид модели, описывающей данное явление.

В то время как статистическую проверку гипотез можно рассматривать в качестве попытки систематизировать процедуру принятия решений на фоне варьирующих наблюдений, статистическое оценивание можно считать попыткой систематизировать интуицию исследователя. При статистическом оценивании в предположении определенной математической модели речь идет о том, чтобы по возможности точно определить истинное значение параметра генеральной совокупности исходя из выборочных наблюдений. Примерами статистического оценивания являются предсказание исходов

выборов и оценка величины брака в промышленном производстве. Если соблюдены предпосылки выборочного метода и принята определенная модель, то можно приступать к построению доверительных границ, определяя интервал, который, например, в 99 из 100 случаев покрывает истинное значение параметра генеральной совокупности.

В попытке систематизировать интуицию факторный анализ представляет шаг вперед по сравнению с методом статистического оценивания. Факторный анализ является методикой, которая в определенном смысле сама является источником возникновения гипотез. Рассмотрим сначала специфический характер гипотез, порождаемых факторным анализом. Исходной предпосылкой служит то, что несколько измеряемых переменных сильно коррелируют между собой. Это означает, что либо они взаимно определяют друг друга, либо связь между этими переменными обуславливается какой-то третьей величиной, которую непосредственно измерить нельзя. Модель факторного анализа всегда связана с последним предположением, т.е. измеряемая величина является лишь формой проявления величины, остающейся на заднем плане и не поддающейся непосредственному измерению. Это предположение во многих случаях вполне реально. Возникает задача, можно ли по данным переменным выделить величину, так называемый *фактор*, который объяснил бы наблюдаемые связи. Слово *фактор* используется в другом смысле, чем это принято в отечественной экономической литературе: речь идет о математической величине, получаемой на основе наблюдений. Спрашивается, какова самая простая линейная гипотеза о структуре этой величины, скрывающейся за коррелированными переменными. Сколько нужно гипотетических величин, или факторов, и каких, чтобы по возможности наиболее точно воспроизвести и объяснить наблюдаемые связи между переменными? Каким образом свести большое количество данных к возможно более простой концепции с минимальной потерей информации?

Получить ответы на такие вопросы — сокровенная давняя мечта исследователей. Факторный анализ пытается это сделать. Образно говоря, он заглядывает за кулисы того, что непосредственно измеряется, и стремится определить истинные функциональные величины, лежащие в основе данного явления. Основная цель факторного анализа состоит в выявлении гипотетических величин, или факторов, по большому числу экспериментальных данных. Факторы должны быть по возможности простыми и достаточно точно описывать и объяснять наблюдаемые величины. Итак, факторный анализ является методом, упорядочивающим кажущуюся хаотичность изучаемого явления и генерирующим новые гипотезы. Число выделяемых факторов должно быть меньше набора исходных величин, структура этих факторов и их взаимосвязь должны быть по возможности более простыми. Возникает вопрос: что же находится за кулисами экспериментальных данных и как может выглядеть система выделенных факторов в самом простом случае?

Исходной предпосылкой анализа является наличие взаимосвязи между несколькими одновременно наблюдаемыми переменными. В качестве количественной меры связи между двумя переменными используется коэффициент корреляции. Он может принимать значения от -1 до $+1$. При этом, если он приближается к нулю, то это свидетельствует об отсутствии линейной связи, и чем более он близок к $+1$ или к -1 , тем более тесная линейная связь существует между переменными. Все вычисленные коэффициенты корреляции располагаются соответствующим образом в корреляционной матрице. В этой матрице содержится важная информация о взаимоотношениях переменных с учетом влияния помех, причиной которых может являться, например, неоднородность материала. При анализе такой корреляционной матрицы получают гипотетические величины, так называемые факторы, которые находятся в определенных взаимоотношениях с переменными. Примечательно то, что факторный анализ делает возможным выдвижение дифференцированных гипотез о

структуре взаимосвязи переменных и факторов, не задаваясь этой структурой заранее и не имея о ней никаких сведений. Эта структура находится по результатам наблюдений. Полученные гипотезы могут быть проверены в ходе дальнейших экспериментов.

Итак, задачей факторного анализа является нахождение простой структуры, которая бы достаточно точно отражала и воспроизводила реальные, существующие в природе зависимости. При этом на реализацию принципа простой структуры оказывают сильное влияние вид и количество экспериментальных данных. Односторонний подход к подбору переменных неизбежно приведет к таким факторам, которые не будут соответствовать изучаемому явлению. Результат факторного анализа определяется постановкой всего исследования. Нельзя надеяться на то, что хаотически собранные данные, подвергнутые факторному анализу по имеющейся стандартной программе, дадут осмысленные результаты. Но не исключено, что конечный результат факторного анализа может оказаться неудовлетворительным также в том случае, когда сбор данных проводился целенаправленно, согласно намеченному плану исследования.

Как в любой прикладной науке, здесь следует обращать внимание на различие между математической моделью и реальным содержанием изучаемого явления. Алгебраически вычислительная сторона метода, в которой речь идет только о решении системы уравнений и точности вычислений, является лишь одним аспектом проблемы. Для факторного анализа характерен также статистический подход, применяемый, например, при проверке гипотезы о числе факторов, подлежащих выделению. Наряду с этим существует еще проблема содержательной интерпретации выделенных факторов, что не имеет места при построении математико-статистической модели. Три вышеназванных аспекта – алгебраически вычислительную сторону, статистический подход и интерпретацию факторов – следует учитывать при проведении факторного анализа и разграничивать их.

Применение факторного анализа особенно эффективно в таких областях, где невозможно манипулировать наблюдаемыми переменными. Исследование поведения человека или его физиологических параметров, а также различных общественных явления нельзя проводить классическим способом, поддерживая на одном и том же уровне определенные величины или вводя различные ограничения. Однако для большинства явлений природы характерна совместная вариация признаков, которая в неизменном окружающем мире дает возможность после точного анализа делать выводы, аналогичные результатам классических экспериментов.

Классическим примером применения факторного анализа является теория интеллектуальных возможностей. Можно предположить, что интеллектуальные возможности состоят из большого числа параметров, характеризующих различные способности человека, такие, как память, внимательность, компетентность и т. д. Этим способностям с помощью психологических тестов пытаются дать количественные оценки и определяют их взаимосвязь. Тесты способностей, как правило, характеризуются положительными корреляциями. Спирмэн [2] предположил, что все возможные совокупности корреляций обуславливаются одним генеральным фактором g , влияющим на все переменные. Этот генеральный фактор совпадает с тем, что обычно в обиходе называют смышленостью. После его психометрической изоляции можно определить, какие методы измерения более всего пригодны для его оценки, т.е. вычисляют нагрузки этого фактора по различным тестам. Обширные исследования показали, что одного генерального фактора недостаточно для описания наблюдений над интеллектуальными способностями. Может быть постулирована более сложная гипотеза, допускающая несколько факторов для адекватного статистического описания интеллекта. Это достигается путем введения новых методов измерения, связанных с решением арифметических задач, с проверкой способностей к абстрактному мышлению, с проверкой словарного запаса и т. д. Другими словами,

методы измерения представляют собой набор тестов, применяемый к репрезентативной выборке индивидуумов. Этим самым исследователи получили в руки измерительный инструмент для оценки гипотетической величины интеллекта, расчлененного на отдельные факторы. С помощью факторного анализа было показано, что интеллектуальные возможности человека можно представить в виде эмпирических показателей, а не только в виде слов или понятий. Эти показатели имеют количественные оценки. Понятие интеллекта заменяется рядом факторов, чьи отображения устанавливаются с помощью методов факторного анализа. На основе психологических тестов, выявляющих интеллектуальные возможности человека, можно выделить набор факторов, которые большей частью коррелированы положительно, например факторы пространственного воображения, понимания слов и т.д. С помощью факторного анализа эмпирическим путем проверялись различные психологические теории интеллектуальных возможностей и умственных способностей человека. Такие исследования нашли практическое применение при оказании помощи в выборе профессии, при распределении рабочей силы по имеющимся рабочим местам, в решении проблем воспитания детей в семье и школе, в психотерапии и т. д.

Аналогично исследованию интеллектуальных возможностей с помощью факторного анализа многие ученые занимались изучением индивидуальных особенностей. Среди этих ученых, прежде всего, следует назвать Каттелла [3], который с помощью почти 2000 различных психологических тестов выделял в многочисленных экспериментах от 50 до 60 факторов. Каждый из этих факторов оценивался с помощью набора психологических тестов, примененных к отдельному испытуемому.

Различные теории, которые возникли в результате использования факторного анализа, хорошо согласуются с известными клиническими наблюдениями и делают возможной классификацию людей по их характеру. Например, с помощью набора тестов можно определить, является ли

испытуемый неврастеником или обладает бодрым и детерминированным характером, причем опыт показал, что результаты тестов соответствуют заключениям врачей.

Факторный анализ применяется и в других областях знаний, например, в социологии – для анализа результатов выборов. Но следует отметить, что факторный анализ как научная методика оправдал себя лишь в отдельных направлениях психологии. Аппарат факторного анализа привлекают в первую очередь к исследованию таких явлений, где необходимо выявить наиболее простую структуру по большому числу наблюдаемых переменных. По мере накопления наблюдаемого материала выявить структуру только с помощью логических рассуждений становится труднее и на это затрачивается много времени. Факторный анализ позволяет произвести упорядочение данных, коротко описать взаимоотношения между переменными и получить вспомогательный материал для проверки интуитивных соображений исследователя.

Факторный анализ занял свое особое место среди других эконометрических методов в силу специфики и особенностей факторной модели. Кендэл [4] указывает в качестве исходной позиции при разграничении многомерных методов различия в понятиях «зависимость» и «взаимозависимость» одной или нескольких переменных от остальных. Конкретная постановка вопроса позволяет определить, что является результативной и факторной величиной. Самым простым примером теории, занимающейся исследованием зависимости, является регрессионный анализ. Кендэл сюда же относит дисперсионный и ковариационный анализ. В противоположность этому под взаимозависимостью понимается исследование связей между несколькими переменными без выделения результативных факторных величин. Сюда относится корреляционный и ассоциативный анализ, анализ функциональных связей, компонентный и факторный анализ. Дискриминантный анализ и вычисление канонических корреляций нельзя отнести ни к одному из этих видов анализа.

Так как факторный анализ исходит из коэффициентов корреляции или ковариации, то его следует считать статистическим методом. Однако определение главных компонент корреляционной матрицы является вычислительной абстракцией, которая не распространяется на генеральную совокупность. Но аналогично определение среднего значения и регрессионной прямой тоже является вычислительной процедурой, осуществляемой по определенному алгоритму. Однако такое представление данных через среднее значение и регрессионную прямую является одной из статистических задач.

В многомерном регрессионном анализе задаются целью получить оптимальную оценку зависимой переменной или результирующего показателя исходя из нескольких независимых переменных, которые точнее будет называть регрессорами. Последние могут зависеть друг от друга, т.е. коррелировать, однако в регрессионном анализе одним из основных предположений является то, что ошибки в измерениях регрессоров, как случайных переменных, пренебрежимо малы по сравнению со случайными отклонениями зависимой переменной.

Результирующий показатель оценивается в целях прогнозирования так, чтобы погрешность была по возможности минимальной. В противоположность этому в факторном анализе пытаются определить новые переменные, так называемые факторы, величины которых могут быть оценены для каждого момента времени в будущем. Если в качестве исходных данных используются не временные значения переменных, а величины показателей ряда различных объектов, то величины факторов могут быть найдены для каждого такого объекта. В регрессионном анализе ограничиваются сопоставлением изменений, лежащих на поверхности явлений, в факторном анализе пытаются заглянуть за кулисы явлений, обнаружить основные причины, лежащие в основе этих изменений, и произвести их оценку в каждом отдельном случае. Модель факторного анализа более сложна. В факторном анализе все переменные одинаково важны и все оцениваются, в то время как в регрессионном анализе производится расчленение переменных. По всем переменным определяется

минимально возможное число независимых факторов, которые затем могут быть оценены для всех рассматриваемых объектов.

Оба метода различаются постановкой целей и ни в коем случае не противопоставляются друг другу. Более того, использование подходов и модели факторного анализа в регрессионном анализе может решить ряд существенных проблем последнего.

Существенное различие между обоими методами состоит в том, что факторный анализ не довольствуется данными переменными, а ищет те немногие величины, которые воспроизводят переменные, взаимосвязь между ними и объясняют их.

В то время как в регрессионном анализе предполагается различие между исходными величинами и целевой функцией, т.е. производится разгруппировка переменных, в дисперсионном анализе исходят из данных, уже сгруппированных по различным объектам. Оценка расхождения между двумя средними значениями выполняется с помощью t -критерия. В рамках дисперсионного анализа значительно расширяются возможности проверки гипотезы существенности различия между парами средних. В случае однофакторного комплекса проверяется гипотеза, является ли существенным различие между средними значениями отдельных групп и средним значением всей совокупности. Вид гипотезы зависит от постановки задачи и от исходных данных.

Большим достоинством дисперсионного анализа является возможность исследования влияния сразу нескольких факторов. В многофакторном дисперсионном анализе проверяется гипотеза различия между векторами средних значений.

Термин «фактор» общепринят в дисперсионном анализе и означает наблюдаемую величину, что отличается от толкования этого термина в факторном анализе.

Постановка задачи в факторном анализе существенно отличается от ее постановки в дисперсионном анализе. Задачей дисперсионного анализа

является исследование влияния одного или нескольких факторных признаков и их взаимодействий на результативный признак и оценка этого влияния. При этом отдельные значения результативного признака увязываются с выбранной моделью. Заранее задаются факторы, число которых ограничено, и их уровни.

В факторном анализе переменные исследуются в отношении их взаимосвязи. При этом заранее не постулируется, что наблюдаемые переменные полностью представляют исследуемую область. Напротив, ставится задача определения количества и вида линейно-независимых друг от друга переменных (факторов), которые бы достаточно точно воспроизводили взаимосвязи наблюдаемых данных. Факторный анализ пытается выявить существенные величины, которые определяют вариацию большого числа переменных. В то время как в дисперсионном анализе имеют дело с одной или несколькими переменными, заданными заранее, и по разработанной методике проверяют их влияние. В факторном анализе пытаются через имеющиеся, ничем не ограниченное число наблюдаемых переменных воспроизвести величины полностью их объясняющие, и оценить эти величины для отдельных объектов.

При конкретных исследованиях имеет смысл оба метода комбинировать. В самом начале, при неопределенной еще структуре исследуемой области, полезно провести факторный анализ, чтобы выявить факторы, вызывающие рассеяние, а потом экспериментировать с ними или, например, провести дисперсионный анализ, чтобы доказать их влияние на третьи величины. Кроме того, можно объединить обе концепции в одном эксперименте и при получении оценок факторов.

Выработка правила, позволяющего приписать некоторый элемент к одной из двух групп, является основной задачей дискриминантного анализа. При такой постановке вопроса должны иметься значения m переменных для n объектов, полученные в результате наблюдений. Кроме того должна быть известна принадлежность каждого объекта к одной из двух групп. К какой из двух этих групп теперь относить новые объекты исходя из имеющейся

информации? Ответ на вопрос упирается в подбор такой линейной комбинации переменных, которая бы оптимально разделила обе группы. В геометрических терминах задача сводится к определению положения новой оси координат, удовлетворяющего следующему условию: проекции объектов двух заранее известных групп должны казаться, возможно, более разделенными. Процедура расчетов не ограничивается двумя группами объектов, ее можно распространить на несколько групп в многомерном пространстве.

Существенное отличие от факторного анализа заключается в том, что в дискриминантном анализе определяется разделение точек, причем заранее уже известно, какие точки к какой группе должны относиться. С помощью факторного анализа в принципе такие группы можно найти.

В то время, как дисперсионный и дискриминантный анализ имеют немного общего с факторным анализом, метод главных компонент, или компонентный анализ, тесно связан с факторным анализом. Этот метод позволяет при заданной n -мерной корреляционной матрице найти новую ортогональную n -мерную систему координат и именно так, чтобы максимум полной дисперсии лежал в направлении первой главной оси, а максимум оставшейся дисперсии – в направлении второй главной оси и т.д. Компонентный анализ является важнейшим методом выделения факторов. Однако его следует строго отличать от факторного анализа как такового.

Метод главных компонент заключается в нахождении последовательности ортогональных осей координат, вдоль которых каждый раз в убывающем порядке определяется максимум полной дисперсии. Процедуру можно прекратить в любом месте и, например, выбрать только первые две главные компоненты, которые воспроизводят 80 процентов полной дисперсии. Проекции объектов на главные компонентные оси можно определить точно. Существенное отличие от факторного анализа заключается в том, что диагональные элементы корреляционной матрицы равны единице, а это означает равенство единице значений общностей, и, как следствие, равенство

нулю значений характеристик¹. Таким образом, в модели компонентного анализа характерные факторы отсутствуют, в чем и состоит главное отличие метода главных компонент и факторного анализа.

Единичная дисперсия всех переменных в компонентном анализе рассматривается в совокупности как общая дисперсия. Хотя это приводит к однозначному решению, однако это нереально во всех практических ситуациях. Пожалуй, никогда не проводится анализ переменных, при котором имела бы место гипотеза, что полная дисперсия равна общей, т.е. что она определяется только наблюдаемыми переменными. Лишь в том случае, когда эта гипотеза подтверждается критерием, можно применять модель компонентного анализа.

Главные компоненты можно рассматривать как возможные математические описания измерений полной дисперсии и им нельзя приписать никакого другого эквивалента в действительности. В соответствии с этим, метод главных компонент позволяет определить ортогональную систему координат с определенными сильными свойствами, которой в действительности чрезвычайно редко могут соответствовать какие-либо условия. Это находится в прямом соответствии с тем, что главные компоненты ортогональны и не подлежат дальнейшему вращению.

В противоположность этому модель факторного анализа позволяет отыскать весь комплекс взаимосвязанных дисперсий. При этом вычислительная операция по определению главных компонент может быть использована в качестве одного из этапов, а именно этапа поиска предварительных факторных решений.

В эконометрии, как в других научных дисциплинах, исследователь постоянно сталкивается с проблемой необходимости разработки приемлемой модели, с помощью которой можно объяснять, а, следовательно, и предсказывать различные экономические явления. Одним из главных источников данных, используемых для этих целей, служат временные ряды

¹ Понятия общности и характерности будут даны ниже. Их сумма равны единице.

наблюдений, которые становятся все более доступными для исследователя, и могут относиться к различным временным интервалам (году, кварталу, месяцу ...). Главной особенностью временных рядов наблюдений, отличающей их от других источников данных, является свойство зависимости наблюдений, относящихся к различным периодам времени. Для того чтобы временные ряды использовались эффективно, необходимо применять динамические спецификации как для детерминированной, так и для стохастической составляющей частей модели. В эконометрическом анализе временных рядов принято считать, что модель полностью определена, если задана совместная функция плотности распределения вероятности для ее зависимых переменных.

Одна из целей теории временных рядов состоит в получении стохастических моделей для дискретных рядов во временной области, обладающих максимальной простотой и минимальным числом параметров и при этом адекватно описывающих наблюдения.

Получение таких моделей важно по следующим причинам:

- они могут помочь понять природу системы, генерирующей временные ряды;
- их можно использовать для оптимального прогнозирования будущих значений рядов;
- если исследуются два или несколько связанных временных рядов, можно расширить модели так, чтобы они описывали динамические взаимосвязи между рядами и, следовательно, позволяли оценивать передаточные функции;
- они могут быть использованы для выработки стратегии оптимального управления; эта стратегия указывает, каким образом надо изменять регулируемую переменную для того, чтобы минимизировать возмущение некоторой зависимой переменной.

Возможность оптимального прогнозирования, понимание динамических взаимосвязей и оптимальное управление имеют большое практическое значение. Например, прогнозирование оптимального сбыта нужно для планирования коммерческих операций; модели передаточных функций

необходимы для улучшения проектирования и контроля заводских процессов; стратегии оптимального управления нужны для регулирования важнейших показателей производственных процессов вручную или при помощи участвующей в процессе вычислительной техники.

В отличие от временных рядов в традиционном факторном анализе предполагается независимость наблюдений, относящихся к различным временным периодам, и в этом состоит принципиальное отличие факторного анализа от анализа временных рядов. С другой стороны, в анализе временных рядов до тонкостей рассматриваются варианты корреляции наблюдений, но в стороне остается вопрос о взаимной связи различных переменных многокомпонентного временного ряда. Проблемы построения передаточных функций в практическом аспекте решены лишь для небольшого числа связанных временных рядов, причем связи предполагаются парными.

Модель факторного анализа с некоторым уточнением может быть распространена на временные ряды, т.е. для наблюдений, не являющихся независимыми, причем, ее использование поможет учесть коррелированность компонент временного ряда, и определить причинную структуру взаимосвязей между компонентами.

Факторный анализ был разработан при проведении психометрических исследований. Он возник из потребности узнать что-либо о структуре взаимосвязей большого числа измеримых переменных. В настоящее время он все еще тесно связан с психологией, однако применение его в эконометрических исследованиях может быть еще более успешным, если хотя бы воспроизвести подходы, применяемые в психометрии, к экономическим объектам, не говоря уже о том, чтобы пойти дальше, развивая методологию факторного анализа с учетом специфики экономических процессов.

Целью метода является выделение из большого числа наблюдаемых переменных наиболее простых показателей (факторов), которые бы описывали данный объект изучения, как можно точнее воспроизводили бы данные, полученные в результате наблюдения, и, в определенном смысле, также

объясняли внутренние объективно существующие закономерности. В принципе факторный анализ не ограничивается лишь воспроизведением исходных данных, что, правда, также является его целью. Факторный анализ решает большой объем задач. С его помощью пытаются разработать такую рациональную модель, которая не только воспроизводит исходные данные, но и позволяет получить интерпретируемую систему показателей модели, имеющих разные эквиваленты в действительности. Затем на втором этапе можно эту модель проверить на новых данных.

Первоочередной задачей факторного анализа является не проверка гипотезы, а ее формирование. Этим факторный анализ отличается от других распространенных статистических методов. Он принадлежит к одному из тех методов, которые делают возможным формирование гипотезы на основе большого комплекса данных. При отыскании факторов, наиболее существенно влияющих на исследуемый признак и стоящих за наблюдаемыми переменными, необходимо руководствоваться принципом как можно более простого объяснения природы данного явления. Это стремление к упрощению находит свое выражение в том, что выделяется такое количество факторов, которое минимально необходимо для воспроизводимости наблюдаемых данных. Это дает возможность более надежно интерпретировать каждый фактор. После построения факторной модели она используется для анализа исследуемого явления.

1.2. Модель факторного анализа

Чтобы численно проиллюстрировать вычислительную сторону факторного анализа, приведем небольшой пример. Пусть в результате наблюдений за пятью показателями получены данные, на основании которых вычислена корреляционная матрица R

$$\mathbf{R} = \begin{pmatrix} 1 & -0,56 & 0,48 & -0,16 & 0,08 \\ -0,56 & 1 & -0,42 & 0,14 & -0,07 \\ 0,48 & -0,42 & 1 & -0,12 & 0,06 \\ -0,16 & 0,14 & -0,12 & 1 & -0,02 \\ 0,08 & -0,07 & 0,06 & -0,02 & 1 \end{pmatrix}.$$

Корреляционная матрица является симметричной, т.е. элементы, лежащие выше диагонали, представляют собой зеркальное отражение элементов, лежащих ниже диагонали матрицы. На главной диагонали всегда располагаются единицы, поскольку коэффициент корреляции показателя самого с собой равен единице.

Целью факторного анализа является извлечение на поверхность величины, так называемого фактора, который, по возможности, точнее позволил бы воспроизвести наблюдаемые корреляции с использованием соответствующей процедуры вычислений. Этот фактор и связанная с ним процедура вычислений в начале исследования являются гипотетическими.

Известно, что симметричные матрицы получаются в результате произведения некоторой матрицы на себя же, но транспонированную. Поэтому, логично было бы предположить, что корреляционная матрица может быть получена не только в результате непосредственного вычисления по исходным данным, а именно:

$$\mathbf{R} = \frac{1}{N-1} \mathbf{Z}^T \mathbf{Z},$$

где $\mathbf{Z}$ – матрица N наблюдений за показателями, центрированными и нормированными (т.е. из каждого значения вычтено среднее и эта разность разделена на среднеквадратическое отклонение);

$[\]^T$ – символ транспонирования.

Тот же результат (но как будет видно ниже – почти тот же) может быть получен с помощью перемножения других матриц, в нашем случае:

$$\mathbf{R} = \mathbf{a} \mathbf{a}^T,$$

где матрица-столбец $\mathbf{a}$ имеет вид:

$$\mathbf{a} = \begin{pmatrix} 0,8 \\ -0,7 \\ 0,6 \\ -0,2 \\ 0,1 \end{pmatrix}$$

В самом деле,

$$\mathbf{a}\mathbf{a}^T = \begin{pmatrix} 0,8 \\ -0,7 \\ 0,6 \\ -0,2 \\ 0,1 \end{pmatrix} \begin{pmatrix} 0,8 & -0,7 & 0,6 & -0,2 & 0,1 \end{pmatrix} = \begin{pmatrix} 0,64 & -0,56 & 0,48 & -0,16 & 0,08 \\ -0,56 & 0,49 & -0,42 & 0,14 & -0,7 \\ 0,48 & -0,42 & 0,36 & -0,12 & 0,06 \\ -0,16 & 0,14 & -0,12 & 0,04 & -0,02 \\ 0,08 & -0,07 & 0,06 & -0,02 & 0,01 \end{pmatrix}$$

В последней матрице видно точное воспроизведение всех элементов корреляционной матрицы, за исключением диагональных элементов. Эти значения называются общностями.

Элементы вектора $\mathbf{a}$ позволяют получить коэффициенты корреляции между разными переменными, причем их число значительно меньше числа наблюдений. Формально говоря, фактически вся информация о взаимосвязях показателей заключена в пяти числах – элементах вектора $\mathbf{a}$. Все это дает основание говорить, что за наблюдаемыми корреляциями стоит некоторая невидимая величина, называемая фактором, которая причинно обуславливает эти корреляции.

В факторном анализе чаще всего применяется линейная модель. Это связано, во-первых, с целью объяснить наблюдаемое явление в терминах более простой теории и, во-вторых, с тем, что отход от линейной модели приводит к математическим трудностям, разрешить которые для ряда практических задач бывает достаточно нелегко.

Обычно в статистическом исследовании рассматривается группа объектов, характеризующихся некоторыми общими для них свойствами. В каждом конкретном случае термину «объект» могут соответствовать самые разные элементы: компании, банки, финансовые бумаги и т.д. Измерения этих свойств

(признаков) объектов будем называть значениями их *переменных*. Всюду далее для обозначения общего числа объектов используется буква N для обозначения числа параметров – буква n . Отдельную переменную будем обозначать через X_j . Индекс i используется для обозначения любого объекта или наблюдения; $i = 1, 2, 3, \dots, N$. Значение параметра X_j для объекта i обозначим X_{ij} , обращая внимание на порядок индексов. Частное значение X_{ij} называется *наблюдаемым (выборочным) значением* и соответствует измерению при произвольном начале координат и произвольных единицах измерения.

В литературе по статистике принято для обозначения статистических характеристик выборки (и гипотетических, или скрытых, параметров) использовать буквы греческого алфавита, а для обозначения наблюдаемых значений параметров – буквы латинского алфавита.

Среднее значение для N наблюдений параметра X_j определяется следующим образом

$$\bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij}. \quad (1.2.1)$$

Математическое ожидание этой переменной будем обозначать греческой буквой μ_j . Наблюдаемое значение переменной может быть преобразовано к более удобной форме, если зафиксировать начало координат и единицы измерения. Пусть начало координат соответствует среднему выборочному значению, тогда величина

$$x_{ij} = X_{ij} - \bar{X}_j. \quad (1.2.2)$$

называется отклонением переменной от его среднего значения.

Выборочная дисперсия, а именно ее несмещенная оценка, переменной X_j определяется в соответствии с выражением

$$s_j^2 = \frac{1}{N-1} \sum_{i=1}^N x_{ij}^2. \quad (1.2.3)$$

Истинное, но неизвестное значение этой дисперсии обозначается σ_j^2 . Теперь приняв в качестве единицы измерения среднеквадратическое

отклонение s_j , определим нормированное значение переменной X_j для объекта или наблюдения i :

$$z_{ij} = \frac{x_{ij}}{s_j}. \quad (1.2.4)$$

Множество значений z_{ij} , ($i = 1, 2, \dots, N$) называется стандартной формой задания переменной Z_j . Очевидно, что математическое ожидание нормированной переменной равно нулю, а ее дисперсия равна единице.

Для двух произвольных переменных X_j и X_k можно определить их коэффициент *ковариации* в соответствии с выражением

$$s_{jk} = \frac{1}{N-1} \sum_{i=1}^N x_{ij} x_{ik}, \quad (1.2.5)$$

а соответствующая характеристика истинного распределения обозначается через σ_{jk} . Коэффициент корреляции обозначается через ρ_{jk} , его выборочное значение подсчитывается из выражения

$$r_{jk} = \frac{s_{jk}}{s_j s_k}, \quad (1.2.6)$$

т.е. он равен отношению коэффициента ковариации к произведению среднеквадратических отклонений переменных.

Обычно первым этапом факторного анализа является вычисление коэффициентов корреляции между всеми изучаемыми переменными.

Задача факторного анализа состоит в том, чтобы выразить, параметр Z_j в терминах скрытых гипотетических *факторов*. Простейшей моделью для описания одной переменной в терминах нескольких других может служить линейная модель. Однако и в рамках линейной модели в зависимости от целей анализа возможны различные варианты. Рассмотрим две такие цели: 1) выделение максимальной дисперсии; 2) «наилучшую» аппроксимацию выборочных корреляций.

К. Пирсон [5] предложил эвристический метод сжатия большого массива информации с одновременным выделением максимальной дисперсии. Позднее Г. Хотеллинг [6] развил этот метод, создав метод *главных компонент*, или *компонентный анализ*. Модель компонентного анализа имеет вид:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jn}F_n, \quad (j = 1, 2, \dots, n), \quad (1.2.7)$$

где каждая из наблюдаемых переменных линейно зависит от n некоррелированных между собой новых компонент $F_1, F_2, \dots, F_n$. Важное свойство метода состоит в том, что каждая очередная компонента дает максимально возможный вклад в суммарную дисперсию параметров. Обычно на практике оставляют небольшое число компонент, особенно в том случае, если на их долю приходится достаточно большой процент суммарной дисперсии переменных. Однако для точной аппроксимации корреляций между переменными необходимы все компоненты.

В противоположность подходу, связанному с максимизацией дисперсии, в модели *классического факторного анализа* требуется наилучшим образом аппроксимировать корреляции. Основная модель факторного анализа может быть записана в таком виде:

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + d_jU_j, \quad (j = 1, 2, \dots, n). \quad (1.2.8)$$

Здесь параметр Z_j линейно зависит от m *общих* факторов и характерного фактора (обычно m много меньше n). Общие факторы учитывают корреляции между переменными, характерный фактор учитывает оставшуюся (в том числе и связанную с различными погрешностями) дисперсию. Коэффициенты при факторах часто называют *нагрузками*.

Если быть точными, то следовало бы факторы и коэффициенты при них в выражении (1.2.8) обозначать греческими буквами, так как здесь речь идет о гипотетических характеристиках генеральной совокупности. Однако обозначения (1.2.8) настолько установились в литературе по факторному анализу, что не стоит менять их. Следует помнить только, что одни и те же символы в (1.2.7) и (1.2.8) обозначают разные вещи.

Используя (1.2.8), можно записать выражение для значения j -го параметра у i -го объекта:

$$z_{ij} = \sum_{p=1}^m a_{jp} F_{ip} + d_j U_{ij}, \quad (i = 1, 2, \dots, N; j = 1, 2, \dots, n). \quad (1.2.9)$$

Здесь F_{ip} – значение общего фактора F_p , для объекта или наблюдения под номером i , а каждое m членов суммы $a_{jp} F_{ip}$ отражает степень влияния соответствующего фактора на параметр Z_j ; $d_j U_{ij}$ – остаток, невязка, в выражении для наблюдаемого значения.

Без потери общности можно предположить, что факторы имеют нулевые средние значения и единичные дисперсии (хотя фактически факторы не известны). Кроме того, считают, что характерные факторы независимы как между собой, так и по отношению к общим факторам. В модели (1.2.8) под факторами F_p , понимают случайные величины с некоторым законом распределения, чаще всего предполагаемым нормальным. Из предположения о том, что факторы – нормально распределенные независимые случайные величины, следует, очевидно, предположение о том, что параметры Z_j имеют многомерное нормальное распределение.

Наибольший теоретический интерес вызывает модель (1.2.8), и в ней – основная проблема факторного анализа: оценка $n \cdot m$ нагрузок общих факторов. Для решения этой проблемы были предложены различные методы, рассмотренные в следующем разделе. В этих методах для вычисления факторных нагрузок используются матрицы коэффициентов корреляций между переменными.

Поскольку в факторном анализе переменная описывается линейной комбинацией некоторого небольшого числа других переменных (плюс остаток), то может показаться, что факторный анализ очень похож на регрессионный анализ. Однако в регрессионном анализе набор независимых переменных предполагается измеряемым в действительности, тогда как в факторном

анализе эти переменные являются гипотетическими и лишь могут быть оценены с помощью измеренной информации при последующем анализе.

В соответствии с моделью (1.2.8) дисперсия переменной может быть выражена через факторы. Так, подставив в выражение выборочной дисперсии (1.2.3) значение нормированной величины, определяемой из выражения (1.2.9), получим:

$$s_j^2 = \sum_{l=1}^N z_{lj}^2 = \sum_{p=1}^m a_{jp}^2 \left(\sum_{l=1}^N F_{lp}^2 / (N-1) \right) + d_j^2 \sum_{l=1}^N U_{lj}^2 / (N-1) + 2 \sum_{p=1}^m \sum_{q=1}^m a_{jp} a_{jq} \left(\sum_{l=1}^N F_{lp} F_{lq} / (N-1) \right) + 2d_j \sum_{p=1}^m a_{jp} \left(\sum_{l=1}^N F_{lp} U_{lj} / (N-1) \right). \quad (1.2.10)$$

Поскольку дисперсия переменной, заданной в стандартном виде, равна единице и все переменные (включая факторы) предполагаются заданными в стандартном виде, предыдущее выражение можно переписать в таком виде:

$$s_j^2 = \sum_{p=1}^m a_{jp}^2 + d_j^2 + 2 \sum_{p=1}^m \sum_{q=1}^m a_{jp} a_{jq} r_{F_p F_q} + 2d_j \sum_{p=1}^m a_{jp} r_{F_p U_j}. \quad (1.2.11)$$

Характерные факторы всегда не коррелированы с общими факторами; принять, что общие факторы не коррелируют также и между собой, то выражение (1.2.11) можно упростить:

$$s_j^2 = 1 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + d_j^2. \quad (1.2.12)$$

Члены в правой части представляют доли дисперсии параметра Z_j , приходящуюся на соответствующие факторы. Например, a_{j2}^2 есть вклад фактора F_2 , в дисперсию параметра Z_j . Полный вклад V_p фактора F_p в суммарную дисперсию параметров определяется выражением

$$V_p = \sum_{j=1}^n a_{jp}^2, \quad (p = 1, 2, \dots, m), \quad (1.2.13)$$

а полный вклад всех общих факторов в суммарную дисперсию переменных равен

$$V = \sum_{p=1}^m V_p. \quad (1.2.14)$$

Отношение $\frac{V}{n}$ служит иногда показателем полноты факторизации.

Из выражения для общей дисперсии (1.2.12) следуют два важных понятия факторного анализа. Первое – *общность* переменной Z_j , которая представляет собой сумму квадратов факторных нагрузок

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2, (j = 1, 2, \dots, n), \quad (1.2.15)$$

и второе – *характерность* – вклад характерного фактора в значение дисперсии. Характерность показывает, насколько в общих факторах учтена суммарная дисперсия переменной. Иногда удобно разбивать характерность на две части: ту, что действительно связана со спецификой изучаемых переменных, и ту, что связана с ошибками измерений. Если к данному набору параметров добавить несколько новых, то изучение добавочных коэффициентов корреляции может привести к появлению дополнительных общих факторов. Это означает, что при рассмотрении первоначального набора переменных эти возможные связи данной переменной отражались в той доле характерности, которая связана с действительной спецификой переменной. Эта доля называется *специфичностью* переменной. Оставшаяся доля, связанная с несовершенством измерений, называется *дисперсией* или *ненадежностью* переменной. Дополнение дисперсии ошибки до полной иногда называют «надежностью» переменной. В психологии эту характеристику переменной (отличая ее от дисперсии ошибки) измеряют обычно значением коэффициента корреляции между двумя отдельными реализациями одного и того же теста или между близкими вариантами теста, реализованными параллельно. Если две такие реализации переменной обозначить Z_{j*} и Z_j , тогда их коэффициент корреляции r_{j*j} измеряет надежность переменной.

Если характерный фактор на два аналогичных вида, то модель (1.2.8) приобретает вид

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + b_jS_j + e_jE_j, (j = 1, 2, \dots, n), \quad (1.2.16)$$

где S_j и E_j – соответственно факторы специфичности и ошибки, b_j и e_j – коэффициенты при них.

Поскольку факторы специфичности и ошибки между собой не коррелируют, то для характерности и ее составляющих частей справедливо соотношение

$$d_j^2 = b_j^2 + e_j^2. \quad (1.2.17)$$

Суммарная дисперсия может быть записана в виде:

$$s_j^2 = h_j^2 + b_j^2 + e_j^2 = 1. \quad (1.2.18)$$

Таким образом, суммарная дисперсия состоит из общности (соответствует факторам F) и характерности, или иначе, суммарная дисперсия переменной состоит из общности, специфичности (соответствует факторам S) и дисперсии ошибки (соответствует факторам E).

Методы факторного анализа позволяют получить для каждой переменной значения общности h_j^2 и характерности d_j^2 . Далее характерность может быть разбита на свои составляющие: специфичность и дисперсию ошибки. При этом способ разбиения не связан с видом факторного решения. К примеру, если из опыта известно значение надежности переменной Z_j , то значение дисперсии ошибки можно получить из соотношения

$$e_j^2 = 1 - r_{j*j}. \quad (1.2.19)$$

Зная дисперсию ошибки, из (1.2.17) можно получить значение специфичности:

$$b_j^2 = d_j^2 - e_j^2, \quad (1.2.20)$$

где характерность d_j^2 известна из результатов факторного анализа. Поскольку надежность является дополнением дисперсии ошибки до единицы, то из (1.2.18) следует, что

$$r_{j*j} = h_j^2 + b_j^2, \quad (1.2.21)$$

и, следовательно,

$$h_j^2 = r_{j*j} - b_j^2 \leq r_{j*j}. \quad (1.2.22)$$

Другими словами, общность любой переменной не превышает его надежности и равно ей только в случае нулевой специфичности.

Доля общности от дисперсии, учитываемой общими факторами, измеренная в процентах, обычно называется *полнотой факторизации* C_j переменной Z_j . Этот показатель обычно меньше 100. Равенство ста процентам возможно лишь в случае отсутствия специфичности у переменной, что практически встречается крайне редко.

Рассмотрим некоторые проблемы, связанные с факторизацией реальных данных. Пользуясь моделью (1.2.8), можно записать выражения для всех n переменных в виде

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{21}F_2 + \dots + a_{m1}F_m + d_1U_1, \\ Z_2 &= a_{12}F_1 + a_{22}F_2 + \dots + a_{m2}F_m + d_2U_2, \\ &\dots \\ Z_n &= a_{1n}F_1 + a_{2n}F_2 + \dots + a_{mn}F_m + d_nU_n \end{aligned} \quad (1.2.23)$$

Такая система в факторном анализе называется *факторным отображением*.

В отображении (1.2.23) общие факторы F_p ($p = 1, 2, \dots, m$) могут быть как коррелированными, так и некоррелированными. В то же время как характерные факторы U_j ($j = 1, 2, \dots, m$) всегда предполагаются не коррелированными как между собой, так и с общими факторами. В уравнении для каждой конкретной переменной некоторые из коэффициентов могут оказаться равными нулю и, следовательно, эта переменная будет выражена через меньшее чем m число общих факторов. Число общих факторов, через которые выражается данная переменная, называется его *сложностью*.

Факторный анализ позволяет получить не только отображение, но и значения коэффициентов корреляции между переменными и факторами. Таблица таких коэффициентов корреляции называется *факторной структурой*.

Для выполнения полного факторного анализа необходимы и отображение и структура.

Рассмотрим теперь функциональные взаимосвязи между элементами структуры и коэффициентами отображения. Умножив любое из уравнений (1.2.23) на соответствующие факторы, произведя суммирование по всем наблюдениям N и разделив на N , получим:

$$\begin{aligned}
 r_{Z_j F_1} &= a_{j1} + a_{j2} r_{F_1 F_2} + \dots + a_{jp} r_{F_1 F_p} + \dots + a_{jm} r_{F_1 F_m}, \\
 &\dots \\
 r_{Z_j F_2} &= a_{j1} r_{F_p F_1} + a_{j2} r_{F_p F_2} + \dots + a_{jp} + \dots + a_{jm} r_{F_p F_m}, \\
 &\dots \\
 r_{Z_j F_m} &= a_{j1} r_{F_m F_1} + a_{j2} r_{F_m F_2} + \dots + a_{jp} r_{F_m F_p} + \dots + a_{jm}
 \end{aligned} \tag{1.2.24}$$

и

$$r_{Z_j U_j} = d_j. \tag{1.2.25}$$

Из уравнения (1.2.25) видно, что коэффициент корреляции переменной с характерным фактором всегда равен коэффициенту при характерном факторе в соответствующем уравнении отображения. Там, где это не будет вызывать недоразумения, будем понимать под факторной структурой таблицу коэффициентов корреляция параметров только с общими факторами, т.е. таблицу значений $r_{Z_j F_p}$.

Иногда уравнениями (1.2.24) пользуются для оценки элементов структуры, но чаще, когда известны коэффициенты корреляции между переменными и факторами и коэффициенты корреляции между самими факторами. С помощью уравнений (1.2.24) получают значения коэффициентов отображения. Формально (1.2.24) можно рассматривать как n систем линейных уравнений

m -го порядка относительно неизвестных a_{jp} ($j = 1, 2, \dots, n$; $p = 1, 2, \dots, m$) с известными левыми частями. Эти системы можно разрешить относительно неизвестных коэффициентов a_{jp} .

Из (1.2.24) следует, что элементы $r_{z_j F_p}$ структуры, вообще говоря, не равны коэффициентам a_{jp} отображения. В случае, когда общие факторы не коррелированы между собой, т.е. $r_{F_p F_q} = 0$, уравнения сворачиваются к виду

$$r_{z_j F_p} = a_{jp} \quad (j = 1, 2, \dots, n; p = 1, 2, \dots, m) \quad (1.2.26)$$

Таким образом, элементы структуры совпадают с соответствующими коэффициентами отображения только в случае некоррелированных факторов. Это означает, что если анализ ведется в предположении некоррелированности факторов, то полное решение получается непосредственно из факторного отображения, а коэффициенты корреляции между переменными и факторами равны коэффициентам отображения.

Как уже указывалось, для выполнения полного факторного анализа необходимы и структура и отображение. Структура позволяет выявить корреляции между переменными и факторами, которые необходимы для идентификации факторов и их последующей оценки. Отображение в форме уравнений регрессии выражает линейные взаимосвязи переменных и факторов. Оно может также служить для воспроизведения корреляций между переменными при проверке значимости решения. Отображение оказывается также полезным при сравнении разных систем факторов для данного набора параметров.

Рассмотрим вопрос, как совокупность факторов объясняет взаимосвязи между параметрами.

Исходная информация представляет собой выборочные коэффициенты корреляции между параметрами. Как же они соотносятся с коэффициентами корреляции, подсчитанными на основе аппроксимирующей линейной модели? Если коэффициенты корреляции, вычисленные из модели, мало отличаются от выборочных коэффициентов корреляции, то говорят, что модель хорошо описывает экспериментальные данные, в противном случае, модель считается неадекватной. Факторное отображение (1.2.23) позволяет следующим образом получить коэффициент корреляции между двумя параметрами: перемножим

любые два уравнения из (1.2.23), просуммируем по всем объектам N и разделим на N . Помня, что факторы заданы в стандартном виде, запишем выражение для коэффициента корреляции между параметрами Z_j и Z_k ($j, k = 1, 2, \dots, n$)

$$\begin{aligned}
 r'_{jk} = & a_{j1}a_{k1} + a_{j2}a_{k2} + \dots + a_{jm}a_{km} + \\
 & + (a_{j1}a_{k2} + a_{j2}a_{k1})r_{F_1F_2} + \dots + (a_{j1}a_{km} + a_{jm}a_{k1})r_{F_1F_m} + \\
 & + (a_{j2}a_{k3} + a_{j3}a_{k2})r_{F_2F_3} + \dots + (a_{j2}a_{km} + a_{jm}a_{k2})r_{F_2F_m} + \\
 & + \dots + a_{j1}d_k r_{F_1U_k} + a_{k1}d_j r_{F_1U_j} + \dots + a_{jm}d_k r_{F_mU_k} + \\
 & + a_{km}d_j r_{F_mU_j} + d_j d_k r_{U_jU_k},
 \end{aligned} \tag{1.2.27}$$

где вычисленный коэффициент корреляции обозначен r'_{jk} в отличие от выборочного коэффициента корреляции r_{jk} . Эти обозначения используются далее повсюду.

Как указывалось выше, характерные факторы предполагаются не коррелированными как между собой, так и с общими факторами, т.е. $r_{F_pU_j} = r_{F_pU_k} = r_{U_jU_k} = 0$. Если общие факторы также некоррелированы, то уравнение (1.2.27) упрощается: $r_{F_pF_q} = 0$, и поэтому все члены под первой строкой исчезают. Таким образом, в случае некоррелированных общих факторов коэффициент корреляции между любыми двумя переменными получается из факторного отображения с помощью уравнения

$$r'_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \dots + a_{jm}a_{km}, \quad (j \neq k, k = 1, 2, \dots, n). \tag{1.2.28}$$

Здесь попросту записана сумма произведений коэффициентов отображения, соответствующих рассматриваемым переменным. Понятно, что коэффициент корреляции параметра с самим собой равен единице. Для вычисленных коэффициентов корреляции это свойство также имеет место благодаря введению характерных факторов. Действительно, положив в (1.2.27) $k = j$, получим коэффициент корреляции параметра с самим собой и, снова предположив некоррелированность факторов, тотчас увидим, что корреляция параметра с самим собой равна сумме общности и характерности параметра.

Выяснив различие между вычисленными и выборочными коэффициентами корреляции, обратимся теперь к тому, что их связывает. Из-за ошибок, вызванных неточностью эксперимента и представительностью выборки, коэффициенты корреляции, вычисленные из факторного отображения, в более общем виде (1.2.27) или в случае некоррелированных факторов в виде (1.2.28), вообще говоря, не равны выборочным коэффициентам корреляции. Это является следствием общенаучного принципа, состоящего в том, что теоретическая модель всегда проще, нежели действительный объект, и потому естественно ожидать несовпадения данных экспериментальных и вычисленных из теории. В частности, в факторном анализе значения коэффициентов корреляции r'_{jk} вычисленных с помощью линейных комбинаций, могут несколько отличаться от выборочных.

После получения факторного отображения необходимо выяснить, насколько адекватно оно описывает корреляции между переменными. Для этого с помощью отображения получают вычисленные корреляции, которые далее вычитаются из выборочных коэффициентов корреляции. Итоговые разности называют *остаточными коэффициентами корреляции*:

$$\bar{r}_{jk} = r_{jk} - r'_{jk}, \quad (1.2.29)$$

где r_{jk} – выборочный коэффициент корреляции, r'_{jk} – вычисленный коэффициент корреляции, полученный с помощью отображения. В случае некоррелированных общих факторов выражение для остаточных коэффициентов корреляции записывается в виде:

$$\bar{r}_{jk} = r_{jk} - (a_{j1}a_{k1} + a_{j2}a_{k2} + \dots + a_{jm}a_{km}). \quad (1.2.30)$$

Теперь, изучая значения и распределение остатков $\bar{r}_{jk}$ можно судить о степени приближения выборочных коэффициентов корреляции вычисленным. Вообще говоря, значения остатков близки к нулю. Поскольку вычисление остатков связано с устранением влияния общих факторов, это означает устранение взаимосвязей между переменными. Можно, следовательно, ожидать,

что остатки распределены так же, как некоторая выборка равного объема с нулевой корреляцией. Стандартное отклонение такой выборки определяется формулой

$$\begin{aligned}\sigma_{r=0} &= \frac{I}{\sqrt{N-I}}, \\ \sigma_{\bar{r}} &\leq \frac{I}{\sqrt{N}},\end{aligned}\tag{1.2.31}$$

где $\sigma_{\bar{r}}$ есть стандартное отклонение совокупности остатков. Такой показатель качества модели не очень хорош, так как он учитывает только размер выборки, в то время как значения остатков могут также зависеть и от других характеристик выборки, в первую очередь, от числа переменных.

На основании критерия (1.2.31), зная только размер выборки, можно делать, например, следующие заключения. Если $\sigma_{\bar{r}}$ значительно больше, чем

$\frac{I}{\sqrt{N}}$, то какие-то связи между параметрами остались неучтенными и,

следовательно, модель нуждается в модификации. Если, наоборот, $\sigma_{\bar{r}}$

значительно меньше $\frac{I}{\sqrt{N}}$ то, возможно, в модель включены несущественные

связи между переменными. И, наконец, если стандартное отклонение остатков лишь немного меньше отклонения для выборки с нулевой корреляцией, то факторное решение может считаться приемлемым в смысле критерия (1.2.31).

В любой науке наблюдаемое явление может быть описано многими, не противоречащими друг другу способами. Выбор конкретного способа интерпретации зависит от цели работы. Эта произвольность или неопределенность, долгое время бывшая предметом исследования ученых, кратко отмечена Мултаном [7]: «любая группа явлений может быть непротиворечиво описана разными путями, вернее, с помощью бесконечно большого числа путей. Независимо от причин, по которым мы выбираем способ интерпретации, мы можем выбрать любой способ, кажущийся нам наиболее целесообразным. Если бы ученые всегда помнили, сколь много приемлемых

интерпретаций наблюдаемых явлений может быть выдвинуто, то быть может, они отказались бы от ультимативных попыток поиска единственно верных научных теорий».

Задачи факторного анализа точно так же являются неопределенными в том смысле, что для заданного набора переменных и коэффициентов корреляции между ними, коэффициенты факторного отображения могут быть вычислены неоднозначно. Иначе говоря, может быть найдено бесконечное число ортогональных (независимых) систем факторов, адекватно описывающих выборочные коэффициенты корреляции. Это свойство было известно математикам еще со времен первых исследований в факторном анализе. Впервые оно было формально изложено Хотеллингом [8] в применении к методу главных компонент. Много позже Т. Андерсон и Г. Рубин [9] дали краткое доказательство этого факта в терминах матричной алгебры.

Неопределенность модели, т.е. неоднозначность факторных нагрузок a_{jp} , имеет своей причиной то обстоятельство, что факторное решение, определяя m -мерное пространство, содержащее общие факторы, не определяет базиса в этом пространстве, а следовательно, не определяет положения факторов в нем. С этой неопределенностью приходится сталкиваться дважды: на первой этапе, — при поиске какого-либо решения, удовлетворяющего модели в статистическом смысле, и на втором этапе, — при придании этому решению вида, наиболее удобного с точки зрения интерпретации. Большинство вычислительных процедур факторного анализа дают неоднозначное решение для факторных нагрузок, исключение составляет метод главных компонент. В то же время, любое решение может быть найдено в каноническом виде после того, как факторное решение, достаточно хорошо описывающее выборку, найдено. Система факторов подвергается такому вращению, чтобы полученная в итоге система (столь же хорошо описывающая выборку, что и исходная) оказалась более интерпретируемой с точки зрения специалистов соответствующей области.

Запишем некоторые из введенных ранее понятий в компактном матричном обозначении. Возникающие при этом удобства связаны не только с простотой записи: часто свойства, ускользающие в обычной громоздкой записи, выявляются в матричных обозначениях. Выпишем в матричной форме некоторые фундаментальные уравнения для факторных отображений и структур.

Введем матричное обозначение некоторых основных понятий факторного анализа. Так, совокупность n рассматриваемых нормированных переменных представляется в виде вектора (матрицы строки)

$$\mathbf{z} = \langle z_1 \ z_2 \ \dots \ z_n \rangle; \quad (1.2.32)$$

совокупность всех N нормированных наблюдаемых значений всех n переменных – в виде матрицы:

$$\mathbf{Z} = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \dots & \dots & \dots & \dots \\ z_{N1} & z_{N2} & \dots & z_{Nn} \end{pmatrix}. \quad (1.2.33)$$

Аналогично представляются факторы:

$$\mathbf{f} = \langle f_1 \ f_2 \ \dots \ f_m \rangle = \mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \dots & \dots & \dots & \dots \\ f_{N1} & f_{N2} & \dots & f_{Nm} \end{pmatrix}, \quad (1.2.34)$$

$$\mathbf{u} = \langle u_1 \ u_2 \ \dots \ u_n \rangle = \mathbf{U} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ u_{1N} & u_{2N} & \dots & u_{Nn} \end{pmatrix}. \quad (1.2.35)$$

Коэффициенты при факторах в факторном отображении можно представить в виде матрицы:

$$\mathbf{M} = \left(\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1m} & d_1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & d_2 & \dots & 0 \\ & & \dots & & & & \dots & \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & \dots & d_n \end{array} \right) = \langle \mathbf{A} \ \mathbf{D} \rangle, \quad (1.2.36)$$

где полная матрица M отображения состоит из матрицы A коэффициентов при общих факторах и диагональной матрицы D коэффициентов при характерных факторах.

Пользуясь введенными матрицами, можно переписать выражение для факторного отображения (1.23) в виде:

$$Z = \begin{pmatrix} f \\ u \end{pmatrix} M^T = \begin{pmatrix} f \\ u \end{pmatrix} \begin{pmatrix} A^T \\ D \end{pmatrix} = fA^T + uD, \quad (1.2.37)$$

Кроме отображения, факторный анализ изучает также структуру, т.е. таблицу коэффициентов корреляции между переменными и факторами (отображение совпадает со структурой только в случае некоррелированных факторов). Коэффициенты корреляции параметров с характерными факторами совпадают с коэффициентами при характерных факторах в (1.2.25), и поэтому диагональную матрицу значений этих коэффициентов корреляции естественно обозначить через D . Факторную структуру можно изобразить в виде матрицы:

$$S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{m1} & s_{m2} & \dots & s_{mn} \end{pmatrix}, \quad (1.2.38)$$

где через s_{pj} обозначены коэффициенты корреляции переменных с общими факторами: $s_{pj} = r_{F_p Z_j}$, ($p = 1, 2, \dots, m; j = 1, 2, \dots, n$).

Выразив в матричной форме факторное отображение и структуру, можно рассмотреть некоторые виды взаимосвязи между ними. Запишем факторное отображение (1.37) с учетом N объектов:

$$Z = FA^T + UD. \quad (1.2.39)$$

Умножим обе части этого выражения на транспонированную матрицу факторных значений и разделим на скаляр $(N-1)$:

$$\frac{1}{N-1} F^T Z = \left(\frac{1}{N-1} F^T F \right) A^T + \left(\frac{1}{N-1} F^T U \right) D. \quad (1.2.40)$$

Левая часть полученного выражения есть не что иное, как $\mathbf{S}$:

$$\frac{I}{N-1} \mathbf{F}^T \mathbf{Z} = \mathbf{S}, \quad (1.2.41)$$

и каждый элемент этой матрицы есть коэффициент корреляции между переменными и факторами. Поскольку факторы как общие, так и характерные заданы в стандартном виде, то выражения в круглых скобках правой части (1.2.40) представляют собой также матрицы корреляций, соответственно, между общими факторами и между общими и характерными факторами. Обозначим матрицу корреляций между общими факторами через $\mathbf{R}_{F,F}$, она имеет вид:

$$\mathbf{R}_{F,F} = \frac{I}{N-1} \mathbf{F}^T \mathbf{F} = \begin{pmatrix} 1 & r_{F_1 F_2} & \dots & r_{F_1 F_m} \\ r_{F_2 F_1} & 1 & \dots & r_{F_2 F_m} \\ & & \dots & \\ r_{F_m F_1} & r_{F_m F_2} & \dots & r_{F_m F_m} \end{pmatrix}. \quad (1.2.42)$$

Одним из предположений факторного анализа есть отсутствие корреляции между общими и характерными факторами, т.е. матрица корреляций между общими и характерными факторами равна нулю:

$$\mathbf{R}_{F,U} = \frac{I}{N-1} \mathbf{F}^T \mathbf{U} = \mathbf{0}. \quad (1.2.43)$$

Подставляя в (40) выражения (41), (42) и (43), получим

$$\mathbf{S} = \mathbf{R}_{F,F} \mathbf{A}^T. \quad (1.2.44)$$

Таким образом, получено фундаментальное соотношение, связывающее факторное отображение $\mathbf{A}$ с факторной структурой $\mathbf{S}$, а именно: матрица факторной структуры равна произведению от умножения матрицы отображения на матрицу коэффициентов корреляции между факторами. Из (1.2.44) видно, что если факторы некоррелированы (т.е. $\mathbf{R}_{F,F}$ есть единичная матрица), то элементы структуры равны соответствующим элементам отображения. Умножив справа обе части (1.2.44) на обратную матрицу корреляций между общими факторами, получим соответствующее выражение для отображения:

$$A^T = R_{F,F}^{-1} S \text{ или } A = S^T R_{F,F}^{-1}. \quad (1.2.45)$$

Пользуясь соотношением между отображением и структурой, можно теперь установить несколько соотношений для матрицы вычисленных корреляций. По определению, матрица выборочных коэффициентов корреляции равна:

$$R = \frac{1}{N-1} Z^T Z. \quad (1.2.46)$$

Подставив в это уравнение (1.2.39), получим:

$$\begin{aligned} R &= \frac{1}{N-1} (FA^T + UD^T)(FA^T + UD^T) = \frac{1}{N-1} (FA^T FA^T + FA^T UD^T + UD^T FA^T + UD^T UD^T) \\ &= \frac{1}{N-1} (AR_{F,F}A^T + AR_{F,U}D + DR_{U,F}A^T + DR_{U,U}D); \end{aligned} \quad (1.2.47)$$

где учтена нулевая корреляция между общими и характерными факторами, $R_{U,U}$ – матрица корреляции между характерными факторами, поскольку они независимы друг от друга и представлены в стандартном виде, то эта матрица равна единичной. Таким образом, выражение (1.2.47) можно записать в виде:

$$R = AR_{F,F}A^T + D^2. \quad (1.2.48)$$

Если общие факторы не коррелируют друг с другом, что в большинстве случаев предполагается в факторном анализе экономических процессов, то матрица корреляции общих факторов равна единичной, и выражение (1.2.48) окончательно приобретает вид:

$$R = AA^T + D^2. \quad (1.2.49)$$

Уравнение (1.2.49) в факторном анализе называется «фундаментальной факторной теоремой». Следует отметить, что правая часть (1.2.49) представляет собой истинные, но неизвестные значения коэффициентов парной корреляции между переменными, которые обозначаются греческой буквой Φ , и, строго говоря, именно эту букву следует употреблять в последнем выражении. Но в факторном анализе настолько «устоялось» употребление обозначений, используемых в (1.2.49), что вслед за этим также будут употребляться эти обозначения. Но там, где будет принципиально важно

подчеркнуть, что речь идет собственно об истинных корреляциях, будет употребляться греческая буква.

Факторный анализ занимается описанием массивов экспериментальных данных (выборочных коэффициентов корреляции) с помощью модели – факторного отображения (1.2.23) или в матричном обозначении (1.2.39). Как видно из (1.2.49), имея факторное отображение, можно вычислить коэффициенты корреляции между переменными, пользуясь только коэффициентами при общих факторах. Для того, чтобы матрица вычисленных коэффициентов корреляции адекватно описывала матрицу выборочных коэффициентов корреляции, необходимо, чтобы ее диагональные элементы вычислялись, исходя из части отображения, отвечающей общим факторам. Таким образом, если на главной диагонали матрицы выборочных коэффициентов корреляции стоят числа, приближающие общности, то факторное решение будет включать и характерные факторы. С другой стороны, если на главной диагонали матрицы выборочных коэффициентов корреляции стоят единицы, то, для того чтобы уравнение (1.2.49) позволило получать единицы в матрице вычисленных коэффициентов корреляции, необходимо, чтобы факторное решение включало только общие факторы. В этой случае не делается предположения о наличии характерных факторов и, следовательно, используется модель компонентного анализа. Если же на диагонали стоят надежности – величины в диапазоне между общностями и единицей, то факторное решение будет включать общие факторы и факторы ошибки, а специфичность будет учтена дисперсией общих факторов. В зависимости от того, какие числа стоят на главной диагонали матрицы выборочных коэффициентов корреляции, общие факторы учитывают разные части дисперсии переменных.

1.3. Методы оценок параметров факторной модели

Прежде чем перейти к рассмотрению новых подходов в использовании методов факторного анализа и оценивании параметров факторной модели, сделаем краткий обзор существующих идей и методов с тем, чтобы принять некоторые простые, но фундаментальные критерии и установить, насколько адекватны этим критериям наиболее употребительные процедуры. Целью настоящей главы является конспективное изложение самых популярных методов факторного анализа. Будут кратко рассмотрены их свойства и указаны существующие между ними различия.

Факторному анализу присуща некоторая неопределенность: корреляционная матрица может быть факторизована бесконечным числом способов. В период наиболее интенсивных исследований этот факт привел к постановке острого вопроса: как находить такое факторное решение, которое в одинаковой степени удовлетворяло бы разных исследователей? Метод главных факторов дает единственное решение, но его конечный результат обычно не устраивает специалистов (по крайней мере, психологов). Поиск «психологически содержательного» решения, которое получалось бы из некоторого произвольного факторного решения с помощью вращения, привел Л. Тэрстоуна к *принципу простой структуры* [10].

Прежде чем перечислить подходы, предложенные для выбора конкретного вида решения (а, следовательно, и метода факторного анализа), рассмотрим в геометрических терминах, в чем состоит эта задача. Набор переменных можно представить множеством векторов (каждый из которых соответствует своей переменной) в пространстве, размерность которого равна числу объектов. В факторном анализе утверждается, что эти векторы, вообще говоря, лежат в пространстве, размерность которого меньше числа переменных. Осями координат этого редуцированного пространства являются общие факторы: исходные переменные могут быть линейно выражены через эти факторы. Задача выявления пространства общих факторов никак не связана с

используемой системой координат. Геометрически факторная неопределенность означает бесконечное число вращений, с помощью которых одна система координат может быть приведена к другой.

Для того, чтобы упростить математическое описание (а иногда и для облегчения содержательной интерпретации), систему координат обычно, так или иначе, меняют. Следует отметить, что при подобных преобразованиях геометрические объекты, например прямая линия или множество точек, не изменяются. Формулы, описывающие объекты могут существенно измениться, но сами объекты относительно такого преобразования, инвариантны.

Обычно исследователь имеет дело только с геометрическими объектами, система координат для него — средство исследования, и он предпочитает одну систему другой, если замена приводит к более простому (или более изящному) описанию объекта. Например, если объект — эллипс — в одной системе координат описывается уравнением, содержащим шесть членов, то можно так подобрать систему координат, что выражение для того же объекта будет выглядеть намного проще, а именно, содержать всего три члена.

В отличие от математика, целью психолога является интерпретация системы координат, а объекты (в частности, множество точек) ему нужны лишь для того, чтобы эту систему координат получить. Рассматриваемые факторным анализом объекты образуют облака точек, каждой из точек соответствует какой-либо тест. Плотность точек, их взаимная близость прямо связаны с корреляциями между тестами. В психологических исследованиях искомая система координат получается на основании конкретных объектов, но упор всегда делается на интерпретацию найденной системы координат.

Критерии, приведенные ниже, имеют целью создать базу для целенаправленного поиска системы координат. В зависимости от выбранного критерия и от того, насколько ему удовлетворяет эмпирическая информация, может быть получено одно из решений, описанных далее. Эти решения названы прямыми факторными решениями, так как они основаны на непосредственном анализе корреляционной матрицы. В отличие от них, существуют косвенные

факторные решения, получаемые разного рода преобразованиями прямых решений.

Для того, чтобы единственным образом описать корреляционную матрицу в терминах системы факторов, необходимо ввести ряд предположений и ограничений. Однако полученное при этом решение будет единственным только с точки зрения принятого критерия. Выбрав другой критерий, получим другое решение, также возможно, единственное. Термин «единственное решение» отражает лишь тот факт, что два исследователя, руководствующиеся одним и тем же критерием и обрабатывающие один массив информации с помощью одной процедуры, придут к одинаковым результатам. Обсудим несколько идей и понятий, лежащих в основе факторного анализа.

1) Факторная модель. Для описания параметров принимается линейная модель. Именно линейная модель лежит в основе классического факторного анализа – выражение (1.2.39).

2) Принцип экономии. Этот принцип, присущий любой научной теории, состоит в том, что закон или модель не могут быть сложнее, чем эмпирические факты, на которых они основаны. Следовательно, *общих факторов* должно быть меньше, чем изучаемых переменных, и, кроме того, линейные описания переменных должны быть достаточно простыми.

3) Вклад факторов. Методы факторного анализа можно различать по вкладу в дисперсию переменных, который вносят отдельные факторы. По одному из критериев требуется, чтобы вклад каждого последующего фактора в суммарную общность был меньше, чем вклад предыдущего. При другом подходе считают, что вклады всех факторов должны быть примерно равными.

4) Группировка переменных. В ряде методов факторного анализа возникает необходимость в группировке переменных. Иногда для этого требуется грубо разбить переменные на группы, в других случаях нужны более тонкие методы группировки. Примерами такого рода методов могут служить «коэффициенты принадлежности», равные произведениям коэффициентов

факторных нагрузок при соответствующих переменных, и метод классификации, развитый Р. Трайном [11].

Если изучаемые переменные пронумеровать таким образом, чтобы переменные из одной группы шли один за другим, то корреляционная матрица приобретет блочный вид. Причем в частях матрицы, соответствующих коэффициентам корреляции между переменными разных блоков, будут стоять значения близкие к нулевым по сравнению с коэффициентами корреляции внутри группы.

5) Система координат. При выборе метода анализа нужно заранее постулировать, будет ли финальная система факторов ортогональной или косоугольной, т.е. будут ли общие факторы коррелированными или не коррелированными между собой. Обычно выборочные коэффициенты корреляции одинаково хорошо описываются в терминах обеих систем факторов. Если конечной целью исследования является косоугольная система, то все равно удобно начинать с ортогональной системы, переходя от нее к косоугольной с помощью различного рода преобразований.

6) Точечное представление, эллипсоидальное распределение. При геометрическом описании переменных каждый из N объектов представляется в виде точки в n -мерном пространстве, каждая ось которого соответствует одной из переменных. Из геометрической теории [12] известно, что на точки может быть натянуто m -мерное пространство общих факторов. Обычно облако точек имеет форму, более или менее близкую к m -мерному эллипсоиду, и становится в точности эллипсоидом в случае нормального распределения [13]. Естественно поэтому взять в качестве системы координат главные оси этого эллипсоида, соответствующее описание называется эллипсоидальным.

7) Векторное представление, линейное описание. Переменные можно рассматривать как векторы в N -мерном пространстве, при этом коэффициент корреляции между двумя переменными определяется как косинус угла между ними или скалярное произведение проекций этих двух векторов на пространство общих факторов. Группа сильно коррелирующих между собой

переменных образует в N -мерном пространстве «узкий конус», т.е. углы между векторами, представляющих переменные, достаточно острые, чтобы косинусы этих углов были близки к единице. Если за ось координат принять вектор, направление которого совпадает с суммой векторов, входящих в такой конус, то все переменные этой группы будут коррелировать с этим вектором на достаточно высоком уровне. Качество линейного описания измеряется тем, насколько направления представляющих векторов близки к направлению оси координат. Построив по средним направлениям отдельных косинусов несколько осей координат, получим, вообще говоря, косоугольную факторную систему.

8) Векторное представление, описание с помощью плоскостей. Линейное описание можно считать плоскостным в том смысле, что, если векторы близки и осям координат, то они близки также и к плоскостям, образованным парами осей координат. Качество *плоскостного* описания можно измерять степенью близости представляющих векторов к координатным плоскостям. Векторы, «хорошие» в смысле плоскостного описания, могут оказаться «плохими» в смысле линейного описания, т.е. может оказаться, что вектор лежит в плоскости координат, но далеко от осей координат. Таким образом, поиск системы факторов, удовлетворяющих плоскостному описанию, дает большую свободу, чем в случае линейного описания. Если окажется, что все векторы переменных лежат в одной из координатных плоскостей, то это означает, что каждая переменная может быть линейно выражена всего через два общих фактора.

9) Векторное представление, описание с помощью гиперплоскостей. В двух предшествующих пунктах для представления подмножеств векторов использовались пространства одного и двух измерений. Данный подход можно распространить на пространства большего числа измерений вплоть до полного пространства факторов.

При описании с помощью гиперплоскостей, лежащих в m -мерном пространстве общих факторов, предполагается, что представляющие векторы

лежат в пространстве $(m-1)$ или меньшего числа измерений. Если все переменные удовлетворяют этому предположению, то, следовательно, сложность любой из переменных меньше числа общих факторов. Подобное утверждение не является сильным, так как оно справедливо уже в случае, когда сложность каждой из переменной равна $(m-1)$. В то же время это утверждение достаточно ценное, поскольку гиперплоскость есть пространство наименьшей размерности, включающее данную переменную. Иными словами, предполагается, что существуют пространства меньшей размерности, чем полное, в которых содержатся некоторые подмножества переменных. В частности, вектор, лежащий на оси координат, содержится в одномерном пространстве, а соответствующая переменная имеет сложность равную единице. Точно так же переменная, вектор которой лежит в координатной плоскости, имеет сложность равную двум. Таким образом, способы описания, приведенные в пунктах 7 и 8, являются частными случаями настоящего описания. Действительно, если переменные представлены с помощью линейного или плоскостного описаний, они могут быть описаны и гиперплоскостями. Обратное, вообще говоря, неверно.

10) Принцип простой структуры. Интуитивное понятие «простой структуры» впервые предложил Л. Тэрстоун [10]. Он же одним из первых заговорил об объективизации самого понятия и предложил процедуру поиска простой структуры.

Тэрстоун предложил следующие три условия, которым должна удовлетворять простая структура:

- 1) в каждой строке факторной структуры должен быть хотя бы один нуль;
- 2) в каждом столбце факторной структуры должно быть по крайней мере m нулей (m – число общих факторов);
- 3) для каждой пары столбцов можно найти по крайней мере m переменных, для которых элементы факторной структура равны нулю в одном из этих двух столбцов и не равны нулю в другом.

Тэрстоун дает такое определение простой структуры: «Если можно подобрать систему координат таким образом, что каждый из представляющих векторов лежал в одной или более гиперплоскости, то полученная система называется *простой структурой*». Как видно, здесь требуется выполнение лишь первого условия — каждой из строк факторной структуры хотя бы одного нуля. Смысл этого определения лучше всего объясняет сам Тэрстоун [10]: «Мы считаем само собой разумеющимся, что индивидуальные различия в остроте зрения не играют роли при осуществлении шага; тем самым мы предполагаем, что те или иные функции (возможности) мозга необходимы индивидууму при решении не любых, а лишь некоторых задач. В этом и заключается принцип «простой структуры».

Из приведенных выше условий второе и третье были введены для того, чтобы избежать неопределенности и совпадения координатных гиперплоскостей. Чтобы зафиксировать систему координат, отвечающую простой структуре, Тэрстоун увеличил число условий до пяти:

- 1) каждая строка факторной структуры должна содержать хотя бы один нулевой элемент;
- 2) в каждом столбце факторной структуры имеется не менее m нулей (m — число общих факторов);
- 3) для каждой пары столбцов найдется несколько переменных, соответствующие элементы которых в матрице равны нулю в одном столбце и отличны от нуля в другом;
- 4) если количество факторов равно четырем или превышает это число, то достаточно велика доля переменных, имеющих в любой паре столбцов одновременно нулевые элементы;
- 5) для любой пары столбцов найдется мало переменных, соответствующие элементы которых в обоих столбцах отличны от нуля.

Принцип простой структуры можно применить как к косоугольному, так и к ортогональному факторным решениям.

Вообще говоря, вращение осей при поиске простой структуры можно понимать как способ уменьшения сложности переменных. Предельным является случай однофакторного решения, когда сложность любой из переменных равна единице. При работе с реальным материалом получить ортогональное однофакторное решение практически невозможно; если же допустить косоугольное решение, то оно будет сильно отличаться от ортогонального. Следовательно, упомянутый случай лишь предельный, но именно он является целью при реализации в многофакторном анализе принципа простой структуры.

Если многофакторное решение удовлетворяет перечисленным выше пяти условиям, то геометрически в плоскости любых двух факторов имеем такую картину: 1) многие точки лежат вблизи осей (факторов); 2) большое число точек располагается вблизи начала координат; 3) относительно малое число точек располагается вдали от начала координат и от обеих осей. Структуры, все двумерные проекции которых образуют описанную картину, Тэрстоун выделил в отдельную группу, отметив, в частности, что в процессе анализа могут получиться диаграммы, которые более, чем другие факты, свидетельствуют о наличии гиперплоскостей простой структуры, а также о том, что полученная структура устойчива и интерпретируема. Подобные двумерные диаграммы являются основой для графических методов вращения.

При решении задач факторного анализа может случиться так, что корреляционная матрица аппроксимирована достаточно точно, но найденная система координат не удовлетворяет исследователя в содержательном плане. Впрочем, возможны разные варианты: иногда решение вполне устраивает исследователя, в некоторых случаях оно необходимо лишь как предварительный этап в поиске интерпретируемого решения. Существуют методы, позволяющие получить решение непосредственно из корреляционной матрицы. Различие между этими методами состоит в том, что если в одних априори нужно знать оценки общностей, а в других методах – число общих факторов, определяющих факторную модель.

Рассмотрим наиболее распространенные и традиционно используемые методы факторного анализа.

Метод главных факторов

Развитие вычислительной техники вывело на первое место по широте применения метод главных факторов.

При поиске системы координат методом главных факторов существенно используется понятие об эллипсоидах. Если в качестве факторов взять главные оси эллипсоида, то каждый последующий фактор будет давать меньший вклад в суммарную общность, чем предыдущий. Другими словами, на первый фактор приходится максимально возможная доля суммарной дисперсии; второй фактор учитывает максимум дисперсии в подпространстве, которое получится после исключения первого фактора (т.е. соответствующей оси координат); третий фактор учитывает максимум дисперсии после исключения первых двух и т.д.; последний фактор учитывает оставшуюся долю общности. Если число общих факторов равно m , то факторное отображение в методе главных факторов имеет вид аналогичный (1.2.23)

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{21}F_2 + \dots + a_{m1}F_m + d_1U_1, \\ Z_2 &= a_{12}F_1 + a_{22}F_2 + \dots + a_{m2}F_m + d_2U_2, \\ &\dots \\ Z_n &= a_{1n}F_1 + a_{2n}F_2 + \dots + a_{mn}F_m + d_nU_n \end{aligned} \quad (1.3.1)$$

Таким образом, в данном методе сложность любой из переменных равна числу общих факторов. Однако, по существу, первый фактор является генеральным фактором. Если все коэффициенты корреляционной матрицы положительны, то все коэффициенты при первом факторе также положительны. В то же время для каждого из остальных факторов $F_2, F_3, \dots, F_m$ примерно половина коэффициентов отрицательна, эти факторы, следовательно, являются биполярными. Биполярный фактор ничем особенным от других факторов не отличается, просто некоторые из переменных дают на нем отрицательные проекции. Можно утверждать, что такие переменные имеют отрицательное значение по данному фактору. Так например, если переменные,

характеризующие «надежность экономического субъекта», дают положительные проекции, то переменные с отрицательными проекциями естественно понимать как «неустойчивость». Очевидно, что без изъятия для решения все знаки при коэффициентах можно поменять на противоположные. Такая операция приведет в данном примере к появлению фактора «неустойчивость экономического субъекта» и группы переменных, измеряющих «отрицательную неустойчивость» (т.е. – надежность).

Метод главных факторов не требует, чтобы все элементы корреляционной матрицы были положительными, таким образом, метод пригоден для анализа любых корреляционных матриц.

Центроидный метод

Максимум популярности центроидного метода приходится на середину прошлого века. В настоящее время его используют в особых случаях, и в частности в виде первого приближения факторного решения, которое затем улучшается с помощью современных методов.

Центроидный метод, давая значительную экономию в объеме вычислительных работ, позволяет аппроксимировать систему главных факторов. Как и в этой системе, каждый последующий центроид (фактор) должен брать на себя максимум суммарной дисперсии. Однако в отличие от метода главных факторов, для данного набора переменных центроидный метод не дает единственного решения. Метод лишен и других важных и интересных свойств, присущих методу главных факторов.

Как и в методе главных факторов, в центроидном методе переменные выражаются через факторы в виде (1.3.1). Если все элементы корреляционной матрицы положительны, то решение состоит из одного генерального и биполярных факторов. С учетом некоторых дополнений центроидный метод позволяет исследовать и матрицы с отрицательными элементами

Треугольная декомпозиция

Метод квадратного корня решения системы линейных уравнений дает возможность свести симметрическую матрицу к такой треугольной,

произведение которой на транспонированную равно исходной матрице. Это свойство составляет одну из основных теорем факторного анализа. Иначе говоря, треугольная декомпозиция симметрической матрицы приводит к факторному решению.

Выразим следующим образом n параметров через n (или меньшее число) некоррелированных факторов:

$$\begin{aligned} Z_1 &= a_{11}F_1; \\ Z_2 &= a_{21}F_1 + a_{22}F_2; \\ &\dots \\ Z_n &= a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nn}F_n. \end{aligned} \tag{1.3.2}$$

Очевидно, что в первое уравнение можно поставить любую из переменных, т.е. система (1.3.2) может быть записана многими способами. При полной корреляционной матрице (с единицами на главной диагонали) в декомпозиции (1.3.2) окажется n факторов. Если заменить диагональные элементы значениями общностей и попытаться извлечь n факторов, то ввиду отсутствия положительной определенности в корреляционной матрице, в ходе реализации метода квадратного корня могут появиться комплексные числа.

Рассмотренные три метода дают решения, которые могут быть как конечным результатом, так и начальным этапом для дальнейших преобразований и улучшений решения. Эти три метода требуют в качестве предварительного шага оценки общностей переменных.

Для последующих методов в качестве предварительного этапа требуется оценка числа общих факторов.

Метод максимального правдоподобия

Основанием этого метода является статистический подход: анализу подвергаются разности между выборочными коэффициентами корреляции и гипотетическими коэффициентами корреляции генеральной совокупности. Первые успешные попытки дать факторному анализу прочный статистический фундамент принадлежит Д. Лоули [14]. Основываясь на методе максимального правдоподобия, Лоули по эмпирическому материалу оценил факторные

нагрузки генеральной совокупности. Метод максимального правдоподобия дает возможность, задавшись числом общих факторов, найти факторное решение и значения общностей. Для проверки гипотез о числе факторов в метод включен статистический тест оценки достоверности.

Хотя формальные вопросы были решены уже к началу 40-х годов, практическая реализация метода максимального правдоподобия стала возможна относительно недавно. Даже несмотря на широкое использование вычислительной техники, решение практических задач с помощью метода максимального правдоподобия оставалось мало значительным. Такое положение дел связано с неудовлетворительной сходимостью итерационной процедуры данного метода. Оказалось, что в теоретическом обосновании метода имелись существенные недостатки. В частности, отмечено [15], что распределение элементов выборочной корреляционной матрицы переменных, подчиненное закону распределения Уишарта, имеет несколько иной вид. Соответственно, неточный вид функции максимального правдоподобия в первоначальном варианте данного метода, привел к проблемам со сходимостью процедуры оценки параметров факторной модели. Что привело к известному забвению этого метода, который является единственным в математической статистике с точки зрения получения наилучших оценок неизвестных параметров.

Как и в методе главных факторов, в методе максимального правдоподобия переменные выражаются через факторы в виде (1.3.1). Однако, получаемые здесь факторы уже не берут на себя максимум суммарной дисперсии. Еще одно различие: если метод главных факторов дает единственное решение, то метод максимального правдоподобия позволяет однозначно определить лишь пространство общих факторов и, следовательно, получить множество решений, эквивалентных с точностью до поворота. Т.е. у данного метода в традиционной трактовке возможно бесчисленное множество равнозначных решений. Чтобы исключить эту неопределенность, необходимо ввести дополнительное условие, которое и зафиксирует конкретное решение.

Метод минимальных остатков

В этом методе факторные нагрузки определяются из условия минимизации внедиагональных элементов остаточных корреляционных матриц. Фактически, это означает реализацию одного из принципов факторного анализа – наилучшее воспроизведение выборочных коэффициентов корреляции. В этом смысле метод отличен от метода главных факторов, направленного на минимизацию дисперсии. В то же время для метода для метода минимальных остатков также справедлива система соотношений (1.3.1). Получаемые здесь решения существенно зависят от того, каким числом общих факторов задаются вначале, а значения общностей получаются попутно, в процессе поиска факторного решения. Эти свойства сближают методы минимальных остатков и главных факторов.

Как видно, исходные предпосылки метода минимальных остатков достаточно близки к классическому подходу факторного анализа.

Групповой метод

В отличие от последних двух методов, результаты которых являются промежуточными и используются для последующего вращения и получения многофакторного решения, групповой метод дает окончательное решение. Общим с этими методами является необходимость предварительной оценки числа общих факторов; кроме того, нужны также и оценки значений общностей. Главное в групповом методе – процедура группировки параметров. На основании произвольно или как-либо разумно отобранных групп строятся общие факторы, число их равняется числу групп.

Естественно ожидать, что система факторов, полученная групповым методом, будет косоугольной. Поэтому полный факторный анализ групповым методом должен включать вычисление факторного отображения, факторной структуры и матрицы коэффициентов корреляции между факторами. Если желательно произвести вращение системы факторов, то в случае ортогональной системы задача упрощается. Применив к матрице факторных коэффициентов корреляции и косоугольной факторной структуре метод квадратного корня,

можно достаточно просто выполнить преобразование косоугольного решения в ортогональное, а также вычислить ортогональное или косоугольное факторное отображение.

Простые факторные модели

Эта группа методов также требует предварительной оценки числа факторов. Практическая их ценность в настоящее время близка к нулю, лишь в очень редких специальных задачах они могут оказаться полезными. Но их крайняя простота и очевидная полезность ретроспективного взгляда на предмет требуют рассмотрения и этой группы методов.

Однофакторное решение является в некотором смысле идеальным решением, в которое трудно вместить реальный материал. Это решение линейное и обладает минимальной сложностью. Дело в том, что каждый представляющим вектор лежит в точности на оси координат и измеряет, следовательно, один фактор. Если для данного материала ортогональное однофакторное решение неудовлетворительно, то можно переходить к косоугольному однофакторному решению.

В пору зарождения факторного анализа для оправдания некоторых психологических теорий были предложены очень простые факторные модели. Так была предложена теория, согласно которой интеллектуальные возможности описываются одним генеральным и одним характерным факторами. В некотором смысле эта теория сводится к однофакторной модели, а именно: предполагается, что все параметры можно выразить через один общий фактор. А это и есть простейший вариант однофакторного решения, когда корреляционная матрица приобретает специфический треугольный вид. По двум типам факторов: генеральному и характерному, эта теория называется бифакторной. Факторная модель характерных и общих факторов является обобщением различных моделей, получивших название по числу рассматриваемых в них факторов.

Особенность бифакторной модели состоит в том, что кроме групповых и характерных, она включает один генеральный фактор. Такое дополнение

позволяет освободиться от ограничения, присущего двухфакторному решению: если с помощью ортогонального однофакторного решения можно описать только корреляционную матрицу треугольного вида, то бифакторное решение может описать уже более сложную картину. По сути дела, такая модификация означает перевод от плоскостного описания к линейному. Действительно, если векторы группы переменных лежат в плоскости координат, то каждая переменная характеризует только два соответствующих фактора. Если же векторы лежат в плоскости, образованной осью генерального фактора (F_0) и одной из осей групповых факторов (а не в плоскости двух групповых факторов), нужно говорить о пучке плоскостей, проходящих через ось F_0 .

Многофакторные методы

В отличие от методов, непосредственно оперирующих корреляционной матрицей, многофакторные методы требуют некоторого исходного, предварительного решения. Термин «многофакторный», введенный Тэрстоуном, выделяет группу методов, в которых система m общих факторов получается без предварительной группировки переменных.

Основная отличительная особенность многофакторного решения состоит в отсутствии генерального фактора и в «перекрытии» групповых факторов. Эффект «перекрытия» состоит в том, что разным групповым факторам соответствуют, вообще говоря, пересекающиеся группы переменных. Следовательно, в описание переменной могут войти несколько групповых факторов. Это означает, что в многофакторном отображении сложность переменных может равняться двум и большему числу (отсюда приставка – «много»). Свойство «перекрытия» позволяет описывать с помощью многофакторных систем матрицы произвольного вида.

Многофакторное решение адекватно описанию с помощью гиперплоскостей. Формально это означает, что сложность параметров не превышает $(m-1)$. Задача, однако, состоит в максимальном упрощении решения, т.е. в уменьшении сложности переменных. Следовательно, несмотря на «перекрытия», желательно получить максимальное число нулевых

коэффициентов в линейных выражениях для переменных. Более точно эти требования отражены в принципе простой структуры.

Любое предварительное решение, полученное на основании анализа корреляционной матрицы, можно преобразовать в ортогональное или косоугольное многофакторное решение. Как правило, в таком решении будет много нулевых коэффициентов. В общем же случае в многофакторном решении необязательны ни нулевые коэффициенты, ни «перекрытие» групповых факторов. На заре использования факторного анализа, при отсутствии достаточной вычислительной базы для получения многофакторных решений привлекались графические методы. В настоящее время имеются формальные процедуры, удовлетворяющие принципу простой структуры.

Прежде чем завершить перечень методов факторного анализа, остановимся еще на одном подходе. Некоторые идеи математической статистики позволяют строить факторное отображение, близкое к заданному. Суть подобных процедур состоит в том, что на основании предварительного решения постулируется, какие факторные нагрузки в окончательном решении будут нулевыми, а какие – большими по величине. Таким образом, предварительное решение преобразуется в конечное, не исходя из некоторых критериев (например, принципа простой структуры), а опираясь на конкретные предположения о виде конечного решения. Следует отметить, что бифакторный метод также приводит к заданному (в некотором смысле) факторному решению. Поэтому, хотя простота и специфика бифакторного метода ставят его несколько в сторону от других, более общих методов, его все же следует причислять к группе методов, приводящих к заданному факторному отображению.

1.4 Проблемы факторного анализа

Исследование латентной природы гипотетических величин – факторов, не может не оставить все проблемы решенными. Само естество объекта

исследования не подлежит очевидному и безапелляционному трактованию по одной причине: невидимый механизм взаимодействия различных явлений является тайной для любого, кто пытается прояснить его, в силу его невидимости. Априорные предположения и косвенные измерения в изучении скрытой от глаз стороны экономического явления в малой степени, если утверждать с формальной точки зрения, могут проявить тайность структурных взаимодействий. Именно эта грань факторного анализа является главным источником его проблем. И главная проблема состоит в отсутствии единственности факторного решения: воспроизведение элементов корреляционной матрицы может быть выполнена бесчисленным количеством способов. Решение однородного алгебраического уравнения также имеет бесчисленное множество решения, а нахождение собственных векторов симметричной матрицы, в чем заключается сущность факторного решения методом главных факторов, приводит именно к решению однородного алгебраического уравнения.

Таким образом, первая проблема факторного анализа – поиск однозначного факторного решения, т.е. нахождение такой матрицы факторных нагрузок, которая не только восстанавливала бы корреляции экономических показателей, но и была бы в некотором роде единственной, порождается самим процессом поиска невидимой структуры. Эта проблема, которая в факторном анализе традиционно решается просто нахождением оценок параметров факторной модели, может быть переформулирована как проблема преодоления неоднозначности факторного решения.

Лоули [14] предлагал разрешить данную проблему, наложив на коэффициенты факторных нагрузок некоторые ограничения. В основе итерационной процедуры в соответствии с методом максимального правдоподобия лежит выражение (обозначения см. (1.2.49))

$$A = R - D^2 \hat{D}^{-2} A J^{-1}, \quad (1.4.1)$$

где

$$\mathbf{J} = \mathbf{A}^T \mathbf{D}^{-2} \mathbf{A}. \quad (1.4.2)$$

При этом факторное решение для однозначности находится таким образом, чтобы матрица $\mathbf{J}$ была диагональной. Это условие соответствует требованию о том, чтобы каждый фактор учитывал максимум дисперсии.

Однако, если учесть, что дисперсия любой переменной состоит из двух компонент, и вклад общего фактора не является единственным, то данное ограничение не выглядит вполне логичным. Несмотря на это, сама идея наложения ограничений на параметры с целью решения проблемы однозначности, является конструктивной. И более уместным было бы введение такого ограничения, которое по возможности в полной мере учитывало аспекты решения максимального круга проблем факторного решения. Тогда такое решение можно было бы считать наилучшим.

Другая проблема факторного анализа – проблема общности.

По определению, общность h_j^2 переменной Z_j равна (1.2.15) – сумме квадратов нагрузок общих факторов. Общность является одной из компонент дисперсии данной переменной. Так как дисперсия каждой переменной после процедуры центрирования и нормализации приводится к единице, то общность является долей единичной дисперсии, которая обуславливается общими для нескольких переменных факторами. Общности являются диагональными элементами матрицы $\mathbf{A}\mathbf{A}^T$, дополнение которых характеристиками делает диагональные элементы равными единице, что соответствует единицам у корреляционной матрицы. Поэтому перед выделением факторов, диагональные элементы корреляционной матрицы должны быть заменены значениями общностей, и это составляет проблему общностей.

Необходимость подбора значений общности меньше единицы вытекает из модели факторного анализа. Каждая переменная обладает общностью, которая должна быть меньше единицы. В предельном случае, когда общая дисперсия определяется только общими факторами, значение общности стремиться к единице, а значение специфичности и величины дисперсии,

обусловленной ошибкой, должны быть около нуля. Такой крайний случай, когда дисперсия ошибки и специфичность равны нулю, на практике не встречается, так как погрешности измерения всегда присутствуют в реальных данных в той или иной мере. А значение специфичности только тогда равно нулю, когда одна и та же переменная дважды включается в анализ. В этом случае специфичные факторы обеих одинаковых переменных становятся общим фактором.

Таким образом, общности могут принимать значения от нуля до единицы. При известной корреляционной матрице эту область принятия значений можно ограничить еще больше. Дуайер [16] доказал, что всегда имеет место неравенство

$$R_{j;1,2,\dots,j(\dots n)}^2 \leq h_j^2; \quad (j = 1, 2, \dots, n), \quad (1.4.3)$$

где $R_{j;1,2,\dots,j(\dots n)}$ – коэффициент множественной корреляции j -й переменной с остальными $(n-1)$ переменными. Квадрат этого коэффициента является нижней границей оценки общности. Более того, при неограниченном возрастании числа переменных при постоянном числе факторов значение общности каждой переменной приближается к квадрату коэффициента множественной корреляции, тем самым неравенство (1.4.3) становится равенством.

Верхняя граница значения общности определяется значением коэффициента надежности, что следует из (1.2.18) и (1.2.19). Таким образом, общность переменной не меньше квадрата коэффициента множественной корреляции и не превосходит квадрат коэффициента надежности:

$$R_{j;1,2,\dots,j(\dots n)}^2 \leq h_j^2 \leq r_{j^*j}^2; \quad (j = 1, 2, \dots, n). \quad (1.4.4)$$

К сожалению, коэффициент надежности переменной вычисляется в редких случаях, так как повторные измерения производятся не часто, а при исследовании экономических процессов это и вовсе затруднено.

Определение общности переменной как суммы квадратов нагрузок общих факторов является формальным и не дает однозначного ответа при оценке общности, поскольку оценка факторных нагрузок должна быть произведена

после оценки общностей. Проблема общности является уязвимым местом многофакторного анализа. Мнения отдельных авторов о способах оценки общностей сильно расходятся, и при разработке методов оценки общностей исходят из различных концепций. В итоге разработано довольно много вычислительных процедур, но все они недостаточно теоретически обоснованы.

Наиболее известный подход к оценке общности заключается в следующем: при известных недиагональных элементах корреляционной матрицы R необходимо подобрать такие значения диагональных элементов, чтобы ранг матрицы с новой диагональю был по возможности минимальным. Однако прямые доказательства того, что такой подход приводит к правдоподобным значениям общностей, отсутствуют.

Для решения некоторых практических задач выбор теоретического подхода имеет второстепенное значение, поскольку при большом числе исследуемых переменных вполне достаточно грубых оценок общности, полученных, например из (1.4.4). Но когда рассматривается небольшое число переменных (не более двадцати) точность, с которой оценена общность каждой переменной, влияет на конечный результат решения.

Перспективным в решении этой проблемы выглядит подход, не требующий предварительной оценки общности, но дающий ее оценку в ходе поиска решения.

Следующая проблема факторного анализа появляется на этапе интерпретации полученного решения и состоит во вращении факторной структуры таким образом, чтобы новые значения факторных нагрузок имели большую наглядность. Проблема вращения состоит в отыскании более предпочтительного положения системы координат путем ее вращения. Переменные можно представить в виде векторов, соотношения между которыми не меняются, какой бы ни была система координат. Вращение подбирают таковым, чтобы максимизировать проекции векторов на одни координатные оси, и минимизировать – на другие. В итоге, значения

факторных нагрузок при одних переменных возрастут, при других должны уменьшиться.

Проблема вращения формулируется в геометрических и алгебраических терминах, что привело к двум подходами к ее решению. Первый из них связан с графическим изображением осей, которые проходят через облака скопления точек – значений наблюдений, отложенных в первоначальной системе координат. Новые оси должны обеспечить большую интерпретируемость факторного решения.

В алгебраических терминах проблему вращения можно сформулировать следующим образом. Пусть получено ортогональное факторное решение $R = AA^T + D^2$, тогда существует бесконечно много других значений факторных нагрузок A^* , которые можно соответствующим путем преобразовать друг в друга и в A : $A^* = AT$, причем $AA^T = A^*A^{*T}$. При осуществлении вращения необходимо найти матрицу преобразования T , которая переводит матрицу A в A^* , причем для A^* должны выполняться определенные условия, и она должна удовлетворять определенным критериям. Если четко сформулировать эти условия и критерии, то можно получить однозначное аналитическое решение. Проблема в том и состоит, чтобы добиться четкости в формулировке этих условий и выработать необходимые критерии, в решении практических задач – это всегда представляет значительную сложность.

До сих пор рассматривалась ситуация, когда для каждого наблюдения имелись различные значения наблюдаемых и регистрируемых переменных. Следующая проблема факторного анализа состоит в получении значений факторов для разных наблюдений. Еще более актуальной эта задача выглядит, когда исходные данные представляют собой значения экономических показателей различных экономических субъектов. Тогда восстановление значений факторов может быть использовано для решения задач классификации объектов. До этого момента рассматривались проблемы, связанные с факторными нагрузками переменных, в проблеме измерения

факторов на первый план выходит задача выражения факторов через переменные, т.е. обратная задача.

В модели факторного анализа матрица Z – это матрица стандартизированных переменных, являющихся линейной комбинацией m факторов, и в ней же присутствует матрица значений факторов F . Она имеет размерность $N \times m$ и в каждой строке содержатся значения факторов отдельных наблюдений или отдельных объектов. Поскольку матрица значений стандартизированных переменных имеет математическое ожидание равное нулю и единичную дисперсию, то значения факторов также должны иметь те же характеристики, т.е. нулевое математическое ожидание и дисперсию равную единице.

Если матрица факторных нагрузок определена, то можно перейти к определению значений факторов. Часто для решения ряда практических задач вполне достаточно знание факторных нагрузок и, соответственно, последующие этапы факторного анализа построены только на анализе корреляций между переменными и факторами. Но для полного проведения факторного анализа, выдвижения новых гипотез относительно исследуемого объекта, необходим этап, на котором оцениваются величины факторов. Требование факторного анализа выявить структуру, завуалированную какими-либо внешними проявлениями, лишь тогда считается выполненной, когда найдено соотношение между этой структурой и отдельным объектом или наблюдением. Каждый экономический объект, таким образом, может быть с погрешностями, обусловленными определением факторов, представлен в виде линейной комбинации значений факторов. С помощью этой комбинации, с другой стороны, можно рассчитать значения реальных экономических показателей у каждого объекта исследования.

Матрицу исходных данных, исследуемую с помощью методов факторного анализа, с точки зрения выявления взаимосвязей можно рассматривать двояко: корреляции можно вычислить между переменными, – с одной стороны. Но коэффициенты парной корреляции можно подсчитать,

используя значения экономических показателей для одного момента времени, т.е. между различными моментами времени или, если статистика набирается по различным экономическим субъектам, то – между ними. Таким образом, к одной и той же матрице исходных данных могут быть применены два способа определения корреляций, которые называются «техниками».

Обычная матрица исходных данных содержит наблюдения по n переменным для N наблюдений или объектов. Если вычислить корреляции между переменными, то получится обычная корреляционная матрица, которая традиционно используется для поиска факторного решения. Факторный анализ этой корреляционной матрицы приводит к обычному факторному отображению. При этой технике с помощью факторов выявляется структура переменных, сосредоточенных вокруг тех факторов, которые определяют эти переменные и обуславливают их корреляции. Применение факторного анализа к такой корреляционной матрице называется R техникой.

Но по той же самой матрице можно рассчитать корреляции между наблюдениями или объектами и получить корреляционную матрицу порядка N . В результате факторизации матрицы коэффициентов между наблюдениями, получается система общих и характерных факторов. Эти факторы вызывают взаимосвязь между наблюдениями, которые разделяются на группы, не похожие друг на друга. Внутри групп существует высокая корреляция между наблюдениями. Применение факторного анализа к матрице коэффициентов корреляции между наблюдениями называется техникой Q .

Хотя обе техники используют матрицы, имеющие общие собственные числа и взаимосвязанные собственные векторы, факторный анализ по ним не дает схожих результатов. Более того, эти техники исследуют природу разных вещей: техника R – взаимосвязи между переменными, Q -техника – корреляции между наблюдениями, и может рассматриваться как аппарат исследования автокорреляционных зависимостей.

В традиционных подходах факторного анализа используются корреляции между случайными величинами, которые представляют собой количественные

характеристики объекта исследования. Количественные характеристики меняют свое значение при переходе от одного элемента совокупности к другому, причем сами изменения носят непрерывный характер. Но имеется ряд показателей, которые не обладают таким свойством. Так несопоставимы между собой различия в школьных оценках, или разные ранги банков по какому-либо показателю (величине собственного капитала, например).

Отдельные значения признака можно упорядочить, руководствуясь местом, которое занимает на шкале измерений каждый объект. В зависимости от используемой шкалы оценок различают четыре способа шкалирования [17].

При применении *номинальной шкалы* каждый объект наблюдения относится к определенной группе, обозначаемой соответствующим числом. Коммерческие банки, например, относятся к группе с номером 1, страховые компании – к группе с номером 2 и т.д. каждый субъект кредитно-финансовой системы относится к определенной группе. Последовательность номеров при таком способе шкалирования может быть любой. В качестве меры взаимосвязи двух величин в таком случае служит коэффициент взаимной сопряженности. Если он статистически значим, то делают вывод о том, что частоты в таблице сопряженности распределены независимо от значений признака. Но коэффициент взаимной сопряженности не обладает одним интересным свойством коэффициента парной корреляции, а именно, коэффициент сопряженности не отражает направления связи.

При использовании *порядковой шкалы* оценок речь идет о ранжировании, аналогичном проставлению школьных оценок. При этом соблюдается порядок, например, $1 < 2 < 3 < 4 < 5$. Знак $<$ означает, что объект с данным рангом является лучше другого в каком-то определенном смысле. Путем сопоставления объектов по какому-либо признаку происходит их упорядочение. Для проверки степени соответствия между двумя последовательностями порядковых статистик служит коэффициент корреляции рангов Спирмена [18].

При интервальном шкалировании оценок расстояние на шкале между элементами рассматриваемой совокупности одинаково. Единица измерения не

изменяется при переходе из одной области шкалы в другую. Школьные оценки, например, этим свойством не обладают. При интервальном шкалировании связь между двумя величинами оценивается по обычному коэффициенту корреляции.

Имеется еще один вид шкал – шкала отношений. Эта шкала редко используется в исследованиях, особенно эконометрических, и здесь также в качестве меры взаимосвязи между объектами используется коэффициент парной корреляции.

Часто не все переменные, включенные в анализ, классифицируются по одной и той же шкале. Часть переменных может быть представлена в виде рангов, некоторые в виде альтернативных данных, например, состояние объекта: нормальное функционирование или банкротство. Если в анализе преобладают признаки с количественной вариацией, то вполне обосновано применение в качестве показателя тесноты связи коэффициента корреляции Пирсона. Если же больше признаков с качественной вариацией, то вычисляют матрицу коэффициентов корреляции рангов. При этом часть информации, которая содержится в данных, являющихся результатом количественных измерений, теряется.

Применение факторного анализа к данным, являющихся результатами измерений качественных признаков, пока еще распространено широко, к тому же оно связано еще с рядом нерешенных проблем, однако такое направление анализа является перспективным.

Еще одной проблемой является неоднородность данных. Например, при исследовании остатков денежных средств на счетах клиентов эта проблема может возникнуть, когда анализируются вместе счета юридических лиц и физических. Очевидно, что закономерности движения денежных средств будут разными. Либо исследование поступлений на депозитные счета может привести к этой же проблеме, когда игнорируется социальный статус клиентов: характер движения денежных средств у состоятельных клиентов и других вкладчиков, скажем, – пенсионеров, будет разным.

Неоднородность данных может привести к появлению фактора, обусловленного только этой неоднородностью. Если он совпадет с каким-либо другим фактором, то нагрузки этого фактора увеличатся по сравнению с исходными.

Неоднородность данных изменяет факторное отображение. Однако при больших изменениях в корреляционной матрице в факторном отображении могут произойти лишь незначительные изменения. Поскольку факторный анализ менее чувствителен к влиянию неоднородности, чем отдельные коэффициенты корреляции, потому что неоднородность может появиться в факторном решении как отдельный фактор и его можно исключить. Но в случае, когда фактор неоднородности совпадает с каким-либо действующим фактором, то факторное отображение меняется значительно.

Таким образом, имеются два типа общих факторов: факторы, которые определяются действием связей между переменными, и факторы, причиной которых является неоднородность данных; плюс к этому еще имеются факторы смешанного типа. В зависимости от постановки задачи, неоднородность может рассматриваться как фактор, искажающий результаты исследования и подлежащий исключению, либо, наоборот, как фактор, вводимый специально для того, чтобы проследить изменение факторного решения. В любом случае, неоднородность не является препятствием для проведения факторного анализа. Неоднородность как раз может быть выявлена благодаря факторному анализу и исключена из решения.

Еще одна проблема, которая решается факторным анализом – это проблема классификации. В Q-технике исследуются связи между объектами, а именно, пытаются найти однородные группы объектов, которые более тесно связаны между собой, чем с объектами других групп. Такая постановка задачи приводит к проблеме классификации. При этом возможны два типа задач.

Первая: классификация при полностью описанных классах. Пусть известно о существовании двух или более совокупностей объектов, у которых измерены значения нескольких показателей. Задача заключается в выработке на

основе имеющихся данных правила, позволяющего некоторый объект отнести к соответствующей совокупности, если заранее не известно, к какой совокупности он принадлежит.

Вторая: классификация объектов при неизвестных классах. Задача состоит в разбиении объектов на классы, как можно более отличающиеся друг от друга. При этом практически отсутствует априорная информация о распределении внутри классов. Статистические критерии, позволяющие производить оптимальное разбиение, еще недостаточно хорошо разработаны. И факторный анализ имеет инструменты, позволяющие решить и эту проблему. В экономической сфере эта проблема стоит на достаточно высоком месте. Она возникает и при решении вопроса о мере близости к банкротству отдельного предприятия, и при ранжировании кредитно-финансовых институтов, и при оценке уровня систематического риска у ряда отраслей, а также при решении очень широкого круга задач классификации.

2. Новый подход к решению проблем факторного анализа

2.1. Оценка общности

Разработано много способов предварительных оценок общностей. Один Тэрстоун [10] предложил двенадцать различных способов. Но пока все способы остаются без исчерпывающего теоретического обоснования.

Один из самых распространенных способов оценки общности состоит в выборе наибольшего коэффициента корреляции данной переменной с остальными переменными. В соответствии с этим способом в строке корреляционной матрицы данной переменной выбирается наибольший коэффициент парной корреляции вне зависимости от его знака, и он используется в качестве оценки общности. При этом отрицательный коэффициент берется с положительным знаком. Никаких дополнительных вычислений не требуется, что является достоинством данного метода. Именно

его простота определила широкое распространение метода в факторном анализе.

Однако наибольший коэффициент корреляции в строке корреляционной матрицы не имеет непосредственной связи с общностью. Коэффициент корреляции является случайной величиной и каждое его возможное значение зависит от случайных обстоятельств, в то время как общность является следствием воздействия на значения переменных и на их корреляции с общими факторами. Поэтому значения общностей и их оценки должны отражать наиболее существенные аспекты факторной структуры. Как уже отмечалось, грубая оценка общности мало влияет на точность факторного решения, но лишь при очень большом числе исследуемых признаков. Когда число переменных невелико (хотя как провести грань между понятиями «велико» и «невелико»), то подобная оценка может существенно исказить факторное решение. Такое исследование провел Иберла [19], которое показало, что даже при незначительном изменении общностей, подставляемых на главную диагональ корреляционной матрицы при оценке параметров факторной модели, значения факторных нагрузок меняются значительно. При этом число переменных было не так уж и мало: оно составляло 12 показателей.

Другим по степени распространенности способом оценки общности является использование коэффициента множественной корреляции, используемого в регрессионном анализе, т.е. меры связи одной переменной со всеми остальными переменными.

Теоретические принципы для этого способа основаны на следующих соображениях. Квадрат коэффициента корреляции между переменной и факторами, определяющими его значения и взаимосвязи с другими переменными, равен общности данной переменной. Такой коэффициент корреляции между переменной и общими факторами равен коэффициенту множественной корреляции, но имеющую другую природу по сравнению с аналогичным понятием, используемым вне факторного анализа. И, если объективно подойти к этой проблеме, то следует отметить, что по

определению, коэффициентом множественной корреляции вектора Z с векторами $F_1, F_2, \dots, F_m$ называется величина

$$\min_{\alpha_1, \alpha_2, \dots, \alpha_m} \sqrt{\left(Z - \sum_j \alpha_j F_j \right)^2},$$

где черта сверху означает среднее значение; таким образом, общность переменной равна квадрату коэффициента множественной корреляции между переменной и общими факторами:

$$h_j^2 = R_{Z_j \cdot F_1, F_2, \dots, F_m}^2. \quad (2.1.1)$$

Поскольку в качестве базиса пространства теоретические значения переменных, получаемых по факторному уравнению, также хороши, как и сами факторы [20], то эти переменные можно взять в качестве базиса пространства, в котором можно определить величины коэффициентов множественной корреляции. Поэтому общности можно получить по выражению

$$h_j^2 = R_{Z_j \cdot Z_1^*, \dots, Z_n^*}^2, \quad (2.1.2)$$

где $Z_1^*, \dots, Z_n^*$ – теоретические значения переменных, полученных по факторной модели. Кайзер [21] показал, что если n векторов $Z_1^*, \dots, Z_n^*$ лежат в пространстве общих факторов, то при вычислении квадрата коэффициента множественной корреляции (2.1.2) некоторые из них можно исключить. Исключение переменной Z_j^* из уравнения факторной модели, приводит к выражению

$$h_j^2 = R_{Z_j \cdot Z_1^*, \dots, Z_{j-1}^*, Z_{j+1}^*, \dots, Z_n^*}^2. \quad (2.1.3)$$

Таким образом, общность переменной равна квадрату его коэффициента множественной корреляции с $n-1$ «общими частями» остальных переменных. Очевидная трудность этого подхода состоит в том, что для вычисления искомой общности, необходимо знать значения остальных $n-1$ общностей. Кайзер предложил итеративную процедуру, рассчитанную на использование вычислительных возможностей компьютерной техники, но процедура сходится

лишь для некоторых корреляционных матриц, отсюда он заключил, что метод «...не имеет практической ценности для эмпирических матриц»[21].

Ранее уже отмечалось, что квадрат коэффициента множественной корреляции, вычисленный из уравнения зависимости данной переменной и остальными, может рассматриваться как нижняя граница оценки общности. Сам коэффициент множественной корреляции известен из регрессионного анализа, и он является мерой дисперсии данной переменной, общей с другими переменными исследуемого множества. В то время как общность является мерой дисперсии данной переменной, обусловленной общими для нескольких переменных факторами. Разность между ними является общей дисперсией, которая может быть приписана только данной переменной, поэтому использовать квадрат коэффициента множественной корреляции, вычисляемой по уравнению регрессии, не может быть оправдано с теоретической точки зрения.

Более того, редуцированная корреляционная матрица, т.е. матрица, у которой главная диагональ заменена значениями общности, уже не будет матрицей Грамма. В этом смысле значения квадратов коэффициентов множественной корреляции не являются оценками общности. Тем не менее, несмотря на теоретическую недопустимость, этот способ получил самое широкое распространение в факторном анализе, в силу того, что он давал наиболее точные по сравнению с другими способами оценки общности.

Перейдем к рассмотрению принципиально нового подхода к получению оценок общностей, основанного на свойствах информации, содержащейся в выборочных данных.

Понятие информации интуитивно толкуется как некоторая совокупность сведений, определяющих меру знаний о тех или иных событиях, явлениях, фактах и прочее. Однако такое определение информации ничего не дает для построения количественной теории информации, которая могла быть использована для решения различных задач и не только инженерных, но математических и экономических.

Очевидно, что всякая информация получается в результате некоторого опыта или наблюдения. Можно сказать, что до опыта нельзя однозначно ответить на интересующий нас вопрос, можно лишь высказать ряд предположений. Таким образом, до опыта имеется некоторая неопределенность в соответствующей ситуации, в исходе тех или иных событий. После опыта, т.е. получения информации, можно ответить на вопрос либо однозначно, либо число предположений уменьшится, и, соответственно, уменьшится неопределенность. Таким образом, для оценки количества получаемой информации необходимо найти меру неопределенности той или иной ситуации.

Уменьшение неопределенности в результате опыта можно принять за наиболее общую меру количества получаемой информации. В этом смысле говорят, что информация обратна неопределенности.

В теории информации известно [22], что информация относительно некоторой точки дискретного пространства x_k (точка такого пространства – это элементарное событие), содержащаяся в y_i – точке другого дискретного пространства, сводится к измерению вероятности x_k от ее априорного значения $p(x_k)$ к ее апостериорному значению $p(x_k | y_i)$. Количество информации, содержащееся в событии y_i относительно появления события x_k , определяется как

$$I(x_k, y_i) = \log \frac{p(x_k | y_i)}{p(x_k)}. \quad (2.1.4)$$

Основание логарифма, используемого в (2.1.4), фиксирует величину единицы измерения информации. Чаще всего используют основание 2, в этом случае она называется «бит». Единица информации, измеренная в битах, означает, что вероятность события x_k увеличилась в 2 раза.

Количество информации (2.1.4) обладает очень важным свойством симметрии по отношению к x_k и y_i , т.е.

$$I(x_k, y_i) = I(y_i, x_k). \quad (2.1.5)$$

Таким образом, можно заметить:

$$I(x_k, y_i) = I(y_i, x_k) \leq \begin{cases} \log \frac{1}{p(x_k)}, \\ \log \frac{1}{p(y_i)}. \end{cases} \quad (2.1.6)$$

О величинах правой части (2.1.6) говорят, что они являются количеством собственной информации для x_k и для y_i . Т.е. собственная информация события x_k обозначается и равна

$$I(x_k) = -\log p(x_k). \quad (2.1.7)$$

Таким образом, информация в каком-либо событии измеряется логарифмом величины обратной вероятности его появления.

Дисперсия переменной случайной величины также несет информацию. По определению, дисперсией называется математическое ожидание квадрата отклонения случайной величины от своего математического ожидания. Тем самым, в дисперсии заключена информация о мере колеблемости рассматриваемой величины. Общая и характерная части дисперсии переменной так же являются частями информации, и к ним может быть применен информационный подход к оценке общности и характерности.

Корреляционная матрица парных корреляций между всеми переменными рассматриваемой совокупности показателей, содержит полную информацию обо всех признаках и взаимосвязях между ними. При этом истинные значения всех корреляций, составляющих матрицу $\mathbf{P}$ (ρ_{ij} – ее элементы), – неизвестны. Факторный анализ исходит из того, что матрица истинных значений корреляций выражается через, опять таки, неизвестные – значения факторных нагрузок (фундаментальная теорема факторного анализа):

$$\mathbf{P} = \mathbf{A}\mathbf{A}^T + \mathbf{D}^2. \quad (2.1.8)$$

Исключение из матрицы какой-либо переменной не поменяет оставшиеся коэффициенты корреляции, но изменит количество информации, содержащейся в ней. Если определить это информационное изменение, то может быть получен ключ к оценке общности переменной.

Поскольку информация измеряется логарифмом обратной вероятности, определим вероятности соответствующих событий. Какие это события?

Информация о самой переменной может содержаться в событии, заключающемся в том, что истинная матрица $\mathbf{P}$ равна выборочной корреляционной матрице без коэффициентов корреляции данной переменной с остальными, т.е. они вычеркнуты из матрицы. Другими словами, реализация этого события говорит о том, что на самом деле присутствие данной переменной в анализе явления необязательно, поскольку она никакой полезной информации не несет и на остальные переменные не влияет. Обозначим это событие $A = \{\mathbf{P} = \mathbf{R}_{-j}\}$.

Другое событие может быть следующим: корреляции между всеми переменными отсутствуют, т.е. истинная корреляционная матрица равна единичной, т.е. такой, у которой на главной диагонали стоят единицы, а остальные элементы – нули. В таком событии заключается информация обо всех корреляциях между переменными, если точно выбрать соответствующую вероятность. Это событие назовем B , и его обозначение имеет вид $B = \{\mathbf{P} = \mathbf{I}\}$.

Вероятности этих событий – условные, поскольку все эти события происходят тогда, когда опыт дает конкретную выборочную корреляционную матрицу $\mathbf{R}$. Таким образом, определим вероятности: $p(\mathbf{P} = \mathbf{R}_{-j} | \mathbf{R})$ и $p(\mathbf{P} = \mathbf{I} | \mathbf{R})$.

Чтобы найти вероятности, нужно знать закон распределения случайных величин. В данном случае такими величинами являются коэффициенты выборочной корреляции.

Известна функция совместного распределения элементов выборочной корреляционной матрицы $\mathbf{R}$, знакомая под названием распределения Уишарта. Она имеет вид [15]

$$W(\mathbf{P}) = \gamma_{N,n} |\mathbf{P}|^{-\frac{N-1}{2}} |\mathbf{R}|^{\frac{N-n-2}{2}} e^{-\frac{N-1}{2} \text{tr}(\mathbf{P}^{-1} \mathbf{R})}, \quad (2.1.9)$$

где N – число наблюдений,

n – число переменных,

tr – операция вычисления следа матрицы (след матрицы равен сумме элементов на главной диагонали),

$\gamma_{N,n}$ – коэффициент распределения Уишарта:

$$\gamma_{N,n} = \left(2^{(N+n+1)/2} \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma\left(\frac{N-i+1}{2}\right) \right)^{-1}, \quad (2.1.10)$$

$\Gamma(\cdot)$ – гамма-функция, определяемая выражениями:

$$\Gamma(x) = (x-1)\Gamma(x-1), \quad \Gamma(1) = 1, \quad \Gamma(1/2) = \sqrt{\pi}.$$

Зная это распределение, можно найти искомые вероятности.

Условная вероятность события A вычисляется по формуле

$$p(\mathbf{P} = \mathbf{R}_{-j} | \mathbf{R}) = \frac{p(\mathbf{R}_{-j} \cap \mathbf{R})}{p(\mathbf{R})} = \frac{p(\mathbf{R}_{-j})}{p(\mathbf{R})}, \quad (2.1.11)$$

а условная вероятность события B – равна

$$p(\mathbf{P} = \mathbf{I} | \mathbf{R}) = \frac{p(\mathbf{I} \cap \mathbf{R})}{p(\mathbf{R})} = \frac{p(\mathbf{I})}{p(\mathbf{R})}. \quad (2.1.12)$$

Используя функцию распределения Уишарта, найдем соответствующие вероятности.

$$p(\mathbf{R}_{-j}) = W(\mathbf{R}_{-j}) = \gamma_{N,n} \left| \mathbf{P} \right|^{\frac{N-1}{2}} \left| \mathbf{R}_{-j} \right|^{\frac{N-n-2}{2}} e^{-\frac{N-1}{2} tr \mathbf{P}^{-1} \mathbf{R}_{-j}} \Bigg|_{\mathbf{P}=\mathbf{R}_{-j}}^{\mathbf{P}=\mathbf{R}}, \quad (2.1.13)$$

После подстановки верхнего, нижнего пределов и небольших преобразований, получаем:

$$p(\mathbf{R}_{-j}) = \gamma_{N,n} \left| \mathbf{R}_{-j} \right|^{\frac{n+1}{2}} e^{-\frac{N-1}{2} (n-1)}. \quad (2.1.14)$$

Вероятность того, что истинные корреляции равны выборочным значениям, может быть получена из выражения:

$$p(\mathbf{R}) = W(\mathbf{R}) = \gamma_{N,n} \left| \mathbf{P} \right|^{\frac{N-1}{2}} \left| \mathbf{R} \right|^{\frac{N-n-2}{2}} e^{-\frac{N-1}{2} tr \mathbf{P}^{-1} \mathbf{R}} \Bigg|_{\mathbf{P}=\mathbf{R}}^{\mathbf{P}=\mathbf{R}}, \quad (2.1.15)$$

в окончательном виде (2.15) становится

$$p(\mathbf{R}) = \gamma_{N,n} |\mathbf{R}|^{-\frac{n+1}{2}} e^{-\frac{N-1}{2}n}. \quad (2.1.16)$$

Вероятность отсутствия корреляций между переменными вычисляется на основании выражения и равно:

$$p(\mathbf{I}) = W(\mathbf{I}) = \gamma_{N,n} |\mathbf{P}|^{-\frac{N}{2}} |\mathbf{R}|^{-\frac{N-n}{2}} e^{-\frac{N-1}{2}tr\mathbf{P}^{-1}\mathbf{R}} \bigg|_{\mathbf{P}=\mathbf{I}} = \gamma_{N,n} |\mathbf{R}|^{-\frac{N-n-2}{2}} e^{-\frac{N-1}{2}n}. \quad (2.1.17)$$

Подставляя найденные вероятности в выражения (2.1.11) и (2.1.12) получим, соответственно для условной вероятности отсутствия влияния переменной Z_j при условии, что она имеется в корреляционной матрице:

$$p(\mathbf{P} = \mathbf{R}_{-j} | \mathbf{R}) = |\mathbf{R}|^{(n+1)/2} |\mathbf{R}_{-j}|^{-(n+1)/2} e^{(N-1)/2}, \quad (2.1.18)$$

и для условной вероятности отсутствия корреляции между переменными, при условии, что в реальности она имеет место:

$$p(\mathbf{P} = \mathbf{I} | \mathbf{R}) = |\mathbf{R}|^{(N-1)/2}. \quad (2.1.19)$$

Теперь найдем информацию, содержащуюся в корреляционной матрице о переменной Z_j – она равна логарифму обратной вероятности:

$$\begin{aligned} I(Z_j, \mathbf{R}) &= -\ln \left(|\mathbf{R}|^{(n+1)/2} |\mathbf{R}_{-j}|^{-(n+1)/2} e^{(N-1)/2} \right) = \\ &= -\frac{n+1}{2} \ln |\mathbf{R}| + \frac{n+1}{2} \ln |\mathbf{R}_{-j}| - \frac{N-1}{2}. \end{aligned} \quad (2.1.20)$$

Информация обо всех переменных в корреляционной матрице имеет вид:

$$I(\mathbf{Z}, \mathbf{R}) = -\ln \left(|\mathbf{R}|^{(N-1)/2} \right) = -\frac{N-1}{2} \ln |\mathbf{R}|. \quad (2.1.21)$$

Информация о самой переменной, которая получается на основании вычисления вероятности о том, что данная переменная в корреляциях не участвует, измеряет «характерность» переменной. Именно изменение обобщенной величины корреляции от нуля до реальной величины, полученной за счет данной переменной, несет такую информацию, однако она является относительной. Теперь, если эту информацию исключить из информации о корреляциях всех переменных, то относительная доля такого изменения даст измерение той части дисперсии, которая приходится на данную переменную со

стороны других, что и составляет общность переменной. Таким образом, значение общности переменной определяется на основании выражения:

$$h_j^2 = \frac{I(Z, R) - I(Z_{-j}, R)}{I(Z, R)} = 1 - \frac{I(Z_{-j}, R)}{I(Z, R)}. \quad (2.1.22)$$

Вычитаемая из единицы величина представляет долю информации о корреляции данной переменной, приходящейся на полную информацию, что численно равно характеристике данной переменной. А сумма общности и характеристики как раз и равна единице, что подтверждает внутреннюю логику и непротиворечивость данных рассуждений.

После подстановки в (2.1.22) величин информаций (2.1.20) и (2.1.21), проведя небольшие преобразования, окончательно получим значение общности переменной Z_j :

$$h_j^2 = 1 - \frac{(n+1) \left(\ln|R| - \ln|R_{-j}| \right)}{(N-1) \ln|R|} \approx N+1. \quad (2.1.23)$$

Теперь остается проверить, насколько значение (2.1.23) близко к истинной величине. Пользуясь реальными данными, сделать это непросто, поскольку истинные значения общностей неизвестны. Более того, их значение, равное сумме квадратов коэффициентов корреляций переменной с факторами, не может не зависеть от числа факторов, которое до момента начала исследования экономического явления так же неизвестно. Поэтому проверка может быть только на смоделированном объекте, таком, как, к примеру, в первом разделе.

Будем считать, что для объяснения корреляций между шестью переменными имеется два фактора. Пусть матрица факторных нагрузок имеет вид:

$$\mathbf{A} = \begin{pmatrix} 0,8 & 0,1 \\ 0,7 & 0,4 \\ 0,9 & 0,2 \\ 0,1 & 0,9 \\ 0,3 & 0,7 \\ 0,2 & 0,8 \end{pmatrix}. \quad (2.1.24)$$

Тогда редуцированная матрица коэффициентов корреляций, т.е. такая матрица, у которой на главной диагонали вместо единиц стоят значения общностей, имеет вид

$$\tilde{\mathbf{R}} = \mathbf{A}\mathbf{A}^T = \begin{pmatrix} 0,65 & 0,6 & 0,74 & 0,17 & 0,31 & 0,24 \\ 0,6 & 0,65 & 0,71 & 0,43 & 0,49 & 0,46 \\ 0,74 & 0,71 & 0,85 & 0,27 & 0,41 & 0,34 \\ 0,17 & 0,43 & 0,27 & 0,82 & 0,66 & 0,74 \\ 0,31 & 0,49 & 0,41 & 0,66 & 0,58 & 0,62 \\ 0,24 & 0,46 & 0,34 & 0,74 & 0,62 & 0,68 \end{pmatrix}, \quad (2.1.25)$$

и эти значения общностей будем считать, с точки зрения вычислительной процедуры, истинными или эталонными.

Сравним с этими значениями значения общностей, полученных посредством предлагаемого метода, и признаваемого на настоящий момент наиболее точным способом – квадрата коэффициента множественной корреляции.

Оценку квадрата коэффициента множественной корреляции данной переменной с остальными переменными можно определить лишь на основании уравнения зависимости этой переменной с остальными, т.е. регрессионного уравнения. Но известен другой способ, в соответствии с которым он определяется на основании выражения [23]

$$R_{Z_j, Z_1 \dots Z_{j-1}, Z_{j+1}, \dots, Z_n}^2 = 1 - \frac{1}{r^{jj}}, \quad (2.1.26)$$

где r^{jj} – диагональный элемент обратной матрицы $\mathbf{R}^{-1}$, стоящий на j -ом месте. Оценки значений общностей, полученных в соответствии с этим

способом обозначим $\hat{h}_{R,j}^2$, т.е. $\hat{h}_{R,j}^2 = R_{Z_j, Z_1 \dots Z_{j-1}, Z_{j+1}, \dots, Z_n}^2$. Соответствующие оценки общностей, полученных по матрице (2.1.25), но у которой на главной диагонали уже стоят единицы, как у выборочной матрицы, имеют вид:

$$\hat{h}_R^2 = \begin{pmatrix} 0,565 & 0,589 & 0,663 & 0,625 & 0,521 & 0,591 \end{pmatrix}, \quad (2.1.27)$$

жирный шрифт означает вектор (матрицу-строку) значений.

Вектор оценок значений общностей, полученных с использованием информационного подхода по выражению (2.1.23) равен:

$$\hat{h}^2 = \begin{pmatrix} 0,639 & 0,635 & 0,621 & 0,629 & 0,646 & 0,635 \end{pmatrix}. \quad (2.1.28)$$

«Крышечка» над символом вектора значений означает оценку неизвестных, но истинных величин. При этом сами истинные величины параметров обозначаются без нее. Приведем эти значения общностей для данного примера:

$$h^2 = \begin{pmatrix} 0,65 & 0,65 & 0,85 & 0,82 & 0,58 & 0,68 \end{pmatrix}. \quad (2.1.29)$$

Для того, чтобы избежать субъективности при сравнении разных оценок, воспользуемся стандартной процедурой анализа соответствий между различными случайными величинами, известной в мире под названием ANOVA. Истинные значения любого неизвестного параметра в математической статистике не являются случайными величинами, таковыми есть их оценки. Истинные общности в факторном анализе также считаются величинами детерминированными. Но в нашем примере, когда элементы факторной структуры воспроизводятся искусственно, «истинные значения» получены в результате вычислительной процедуры, за которой в действительности стояли бы реальные выборочные данные. А, как известно, в выборке не может содержаться вся информация о генеральной совокупности, и в силу случайности выборки, полученные эталонные значения так же наделены природой случайности. Поэтому применение статистического подхода к анализу разных случайных величин на предмет их схожести, который предполагает задаваемую наперед величину ошибки в принятии решения, в данном случае уместен.

ANOVA (ANalysis Of VAriance) – широко распространенный метод выявления различий переменных, получаемых либо разным способом, либо в неодинаковые периоды времени [24]. Итоговым результатом метода служит принятие или отвержение, с определенным заранее уровнем значимости, гипотезы об отсутствии различий у рассматриваемых величин.

Сравним значения оценок общностей – квадратов коэффициентов множественной корреляции – с «истинными значениями». Стандартная таблица ANOVA-анализа с применением однофакторной модели имеет вид, приведенный в таблице 2.1.1. Напомним, фактор в дисперсионном анализе имеет иное значение, чем в статистическом факторном анализе. В данном случае для ANOVA фактором служит измерение общности.

Таблица 2.1.1 – Сравнение квадратов коэффициентов множественной корреляции с «истинными значениями»

	Сумма квадратов	Степени свободы	Средние квадраты	F-значение
Фактора	0,04	1	0,04	5,822
Ошибок	0,068	10	0,0068	
Полная	0,108	11	0,0098	

Итоги анализа различий между оценками общности, полученные с помощью информационного подхода, и значений общности, заданных посредством факторных нагрузок, приведены в таблице 2.1.2.

Таблица 2.1.2 – Сравнение «информационных» оценок общности с «истинными значениями»

	Сумма квадратов	Степени свободы	Средние квадраты	F-значение
Фактора	0,015	1	0,015	2,645
Ошибок	0,057	10	0,0057	
Полная	0,072	11	0,0065	

Для проверки гипотезы о тождественности разных групп величин найденные значения F-критерия необходимо сравнить с табличным значением распределения Фишера. Для данных степеней свободы и уровня значимости α (наперед задаваемой ошибки в статистических рассуждениях), традиционно равной 0,05 в эконометрических исследованиях, эта величина имеет значение 4,965. Если вычисленное значение F-критерия больше табличного, то гипотеза об идентичности величин отклоняется с вероятностью ошибки 0,05. Если же F-значение меньше табличного, то гипотеза принимается.

Проверка показала, что у оценок общности, как квадратов коэффициентов множественной корреляции, с эталонными значениями нет ничего общего. И, напротив, предлагаемые оценки общности на основе информационного подхода согласуются с «истинными значениями» общности.

Знание более точных значений оценок общности позволит повысить точность эконометрического исследования с использованием факторного анализа.

2.2. Оценка факторных нагрузок ортогональной факторной модели

Центральной проблемой факторного анализа является проблема оценки параметров факторной структуры. Именно ее решение позволяет продуктивно использовать факторный анализ для решения различных задач. Более того, без оценок факторных нагрузок факторная модель не представляет собой никакой ценности, поскольку именно количественный, а не качественный, анализ составляет основную идею факторного анализа.

В предыдущем разделе указывались основные методы и способы получения оценок параметров факторной модели. Каждый из них имеет свои достоинства, но и каждому присущи свои недостатки. Общая проблема, связывающая все существующие методы, состоит в отсутствии однозначного решения задачи оценивания. Этот недостаток, однако, имеет свою

положительную сторону: используя методы вращения факторного решения, можно получить более интерпретируемое решение. Для получения однозначного решения необходимо сформулировать дополнительные ограничения на параметры факторной структуры, и получить оценки в виде условного экстремума.

Поскольку большинство методов факторного анализа дают оценки параметров в ходе итерационных процедур, сходимость которых в значительной мере зависит от начального приближения, то следует в самом начале исследования получить достаточно хорошее приближение. В качестве начального «хорошего» приближения традиционно используются собственные векторы корреляционной матрицы. Как известно, компонентный анализ в качестве своего решения использует эти векторы. Но в компонентном анализе предполагается наличие такого количества компонент (факторов) сколько переменных рассматривается в анализе. Т.е. на предварительном этапе не учитывается число факторов. Поэтому целесообразно на этапе получения предварительного решения задать число факторов и модифицировать собственные векторы так, чтобы они приблизились к истинным значениям.

Основания для такого подхода достаточно простые: редуцированная матрица получается в результате перемножения матрицы факторных нагрузок:

$$AA^T = \tilde{R} \quad (2.2.1)$$

где $\tilde{R}$ – редуцированная матрица, т.е. матрица у которой на главной диагонали стоят значения общностей.

Редуцированная корреляционная матрица может быть разложена на собственные векторы и собственные значения:

$$\tilde{R} = QLQ^T, \quad (2.2.2)$$

где Q и L – матрицы, соответственно, собственных векторов и собственных значений редуцированной матрицы.

Последнее выражение может быть подставлено в (2.2.1)

$$AA^T = QLQ^T. \quad (2.2.3)$$

После несложных преобразований (2.32) преобразуется к виду

$$A = L^{-1}Q^T A A^T Q A \quad (2.2.4)$$

Выражение (2.2.4) дает основание для итерационной процедуры поиска значений матрицы факторных нагрузок, которые построены на основании собственных значений редуцированной корреляционной матрицы с учетом информации о числе факторов. Процедура будет построена на том предположении, что значение матрицы A , стоящей в левой части выражения (2.2.4) уточняется вычислительной процедурой действий над матрицами, стоящими в правой части этого выражения. Таким образом, простая итерационная процедура задается уравнением

$$A_{i+1} = L^{-1}Q^T A_i A_i^T Q A_i, \quad i = 1, 2, \dots \quad (2.2.5)$$

где i – номер итерации, которая будет продолжаться, пока матрицы двух соседних итераций не будут отличаться друг от друга на величину наперед задаваемой ошибки.

Практика показывает, однако, что процедуры, подобные (2.2.5) имеют проблемы со сходимостью, которое зависит от начального приближения. Итерационная процедура может и не дать сходящийся результат для некоторых исходных данных. Причем теоретические исследования сходимости многомерных операторов известны для небольшого круга матричных итерационных процедур.

Процесс сходимости можно ускорить и, в большинстве случаев, гарантировать, если воспользоваться матричным аналогом метода Ньютона-Канторовича численного решения трансцендентного уравнения. Кратко опишем его.

Пусть необходимо решить матричное уравнение, не имеющее аналитического решения

$$F(X) = 0, \quad (2.2.6)$$

т.е. в ходе итерационной процедуры найти такое значение матрицы X , которое удовлетворяло бы уравнению (2.35). Такая процедура имеет вид

$$I \otimes X_{i+1} = I \otimes X_i - (I \otimes H F(X_i)) \left(\frac{\partial F(X_i)}{\partial X_i} \right)^{-1}, \quad (2.2.7)$$

где $\otimes$ - знак произведения Кронекера,

I - единичная матрица,

i - также как и для (2.34) означает номер итерации,

H - идемпотентная матрица, имеющая вид

$$H = (X_i - X_{i-1})((X_i - X_{i-1})^T (X_i - X_{i-1}))^{-1} (X_i - X_{i-1})^T, \quad (2.2.8)$$

на первом шаге, когда известна только X_0 , в качестве идемпотентной матрицы берется единичная, поскольку она, также как и идемпотентная, обладает свойством:

$$HH = H.$$

Заметим также, что для обращения производной она должна быть квадратной матрицей, что возможно в случае, когда матричный функционал $F(X)$ имеет ту же размерность, что и неизвестная матрица X .

В нашем случае уравнением, подлежащим решению, служит выражение (2.2.4), перепишем его в форме (2.2.6), но неизвестную матрицу будем обозначать A :

$$A - L^{-1} Q^T A A^T Q A = 0, \quad (2.2.9)$$

Левая часть (2.2.9) представляет собой матричный функционал $F(A)$, который имеет ту же размерность, что и неизвестная матрица A . Его производная равна

$$\frac{\partial F(A)}{\partial A} = I_{nm} - 2(QL^{-1} \otimes A^T Q A) - Q A A^T Q L^{-1} \otimes I_m, \quad (2.2.10)$$

где n - число переменных;

m - число факторов

и нижний индекс единичной матрицы указывает на ее размерность.

Тогда итерационная процедура в соответствии с (3.36) задается выражением:

$$I \otimes A_{i+1} = I \otimes A_i - \left(I \otimes H_i (A_i - L^{-1} Q^T A_i A_i^T Q A_i) \right) \cdot \\ \cdot (I_{nm} - 2(QL^{-1} \otimes A^T Q A) - Q A A^T Q L^{-1} \otimes I_m)^{-1}, \quad (2.2.11)$$

где $H_i = (A_i - A_{i-1})((A_i - A_{i-1})^T(A_i - A_{i-1}))^{-1}(A_i - A_{i-1})^T$ для $i = 1, 2, 3, \dots$ и $H_1 = I$.

Вычислительные эксперименты показывают, что сходимость достигается достаточно быстро, причем решение получается всякий раз, каким бы ни было начальное приближение. Поэтому в качестве начальной матрицы A_0 удобно брать часть редуцированной корреляционной матрицы, а именно первые m ее столбцов, или столько же любых других ее столбцов.

После того как получено хорошее приближение факторных нагрузок, их значения можно улучшить, воспользовавшись методом максимального правдоподобия. В математической статистике это наиболее мощный инструмент получения оценок на основании выборочных данных. Суть его состоит в получении таких оценок, которые бы максимизировали функцию правдоподобия. Такой функцией, как правило, служит совместная плотность распределения случайных величин, по которым имеются ряды статистических данных. Как известно [25] совместной плотностью является произведение плотностей распределения отдельных рассматриваемых случайных величин. И в соответствии с методом максимального правдоподобия значение оцениваемого параметра подбирается так, чтобы значение совместной плотности было бы максимальным, а это соответствует максимуму вероятности того, что оценка совпадает с истинным, но неизвестным значением параметра.

Максимум любой функции определяется после приравнивания нулю первой производной, но поскольку дифференцировать произведения, а после чего искать корень уравнения производной нулю, сложно, то часто совместную плотность логарифмируют. И именно логарифм плотности считают функцией правдоподобия, тем более, что функция и ее логарифм имеют максимум в одной точке.

В факторном анализе неизвестным параметром, предназначенным для оценивания, является матрица факторных нагрузок, которая определяет значения истинной корреляционной матрицы. Плотность совместного

распределения коэффициентов корреляций и составляет основу функции правдоподобия. Совместное распределение элементов корреляционной матрицы подчинено закону Уишарта, плотность распределения которого имеет вид (2.1.9). Логарифмирование и исключение из его результата слагаемых и множителей, не зависящих от матрицы факторных нагрузок дает следующее выражение для функции правдоподобия:

$$g = \ln|\mathbf{P}| + \text{tr}(\mathbf{P}^{-1}\mathbf{R}), \quad (2.2.12)$$

где $\mathbf{P}$ – истинная корреляционная матрица, имеющая вид:

$$\mathbf{P} = \mathbf{A}\mathbf{A}^T + \mathbf{D}^2,$$

$\mathbf{D}^2$ – диагональная матрица характеристик каждой переменной;

$\mathbf{R}$ – матрица выборочных корреляций.

Теперь найдем производную функции правдоподобия g по неизвестному параметру $\mathbf{A}$:

$$\frac{\partial g}{\partial \mathbf{A}} = 2\text{vec}(\mathbf{P}^{-1}\mathbf{A}) - 2\text{vec}(\mathbf{P}^{-1}\mathbf{R}\mathbf{P}^{-1}\mathbf{A}), \quad (2.2.13)$$

где vec – операция преобразования матрицы в вектор – матрицу столбец, путем взятия строки за строкой.

Затем приравняем производную, т.е. выражение (2.2.13), нулю, и найдем выражение, дающее основание для итерационной процедуры:

$$\mathbf{A} = \mathbf{R}\mathbf{P}^{-1}\mathbf{A}. \quad (2.2.14)$$

Авторы метода максимального правдоподобия в факторном анализе Лоули и Максвелл [14], а вслед за ними и известный эксперт в области факторного анализа – Харман [23] отмечали сложности в сходимости процедур, построенных на выражении (2.2.14). Ради улучшения сходимости равенство (2.2.14) преобразовывали к виду, который давал возможность трактовать его как способ вычисления собственных векторов некоторой матрицы, упрощали с целью снизить размерность обрабатываемой матрицы (это было особенно актуально во времена ограниченных возможностей вычислительной техники) и тому подобное. Более того, исследование сходимости итерационных процедур метода максимального правдоподобия продолжается и по сей день. Однако

матричный аналог метода Ньютона-Канторовича (2.2.17) позволяет получить гарантированно сходящуюся процедуру получения максимально правдоподобной оценки матрицы факторных нагрузок.

Матричным функционалом, нуль которого необходимо найти, является выражение:

$$F(A) = A - RP^{-1}A. \quad (2.2.15)$$

Для составления итерационной процедуры найдем производную $F(A)$:

$$F'(A) = I_{nm} + P^{-1}R \otimes A^T P^{-1}A + E_{nm}(A^T P^{-1}R \otimes P^{-1}A) - \\ - P^{-1}R \otimes I_m. \quad (2.2.16)$$

где I_{nm} – единичная матрица, размерностью $nm \times nm$;

E_{nm} – перестановочная матрица, возникающая при дифференцировании транспонированной матрицы по исходной, она обладает свойством:

$$E_{nm}^T E_{nm} = E_{nm} E_{nm}^T = I_{nm}.$$

К примеру, пусть матрица A имеет три строки и два столбца, тогда результатом дифференцирования транспонированной матрицы A^T по исходной будет матрица $E_{3,2}$, которая выглядит следующим образом:

$$E_{3,2} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Итерационная процедура состоит из нескольких этапов. На первом этапе оцениваются общности переменных и, соответственно, – характеристики D^2 . Итерационная процедура (2.40) дает оптимальное начальное приближение A_0 для получения максимально правдоподобных оценок матрицы факторных нагрузок. Затем для получения значений идемпотентной матрицы (2.37) на первой итерации определяются значения матриц факторных нагрузок и корреляций:

$$A_1 = A_0 - RP_0^{-1}A_0; \\ P_1 = A_1 A_1^T + D^2.$$

После этого основная итерационная процедура имеет вид:

$$\begin{aligned} I_n \otimes A_{i+1} = I_n \otimes A_i - \left(I_n \otimes H_i (A_{i-1} - R P_{i-1}^{-1} A_{i-1}) \right) \cdot \\ \cdot (I_{nm} + P_i^{-1} R \otimes A_i^T P_i^{-1} A_i + E_{nm} (A_i^T P_i^{-1} R \otimes P_i^{-1} A_i) - \\ - P_i^{-1} R \otimes I_m)^{-1}, \end{aligned} \quad (2.2.17)$$

где $H_i = (A_i - A_{i-1})((A_i - A_{i-1})^T (A_i - A_{i-1}))^{-1} (A_i - A_{i-1})^T$,

$i = 1, 2, \dots$ и процедура выполняется до тех пор, пока матрицы факторных нагрузок соседних итераций не станут отличаться на величину наперед заданной погрешности.

Выделить матрицу A_{i+1} из кронекеровского произведения несложно, но процедуру (2.2.17) можно упростить, если воспользоваться свойством умножения произведения Кронекера на вектор:

$$(A \otimes B) \text{vec}(C) = \text{vec}(ACB^T).$$

Введем матрицу свертки S , имеющей такую же размерность, как и матрица факторных нагрузок A , т.е. $n \times m$. Обе матрицы имеют одинаковый ранг, равный числу факторов m . Матрица свертки – блочная: верхняя ее часть представляет собой диагональную матрицу, а нижняя часть – целиком состоит из нулей. К примеру, матрица факторных нагрузок имеет пять строк и три столбца, тогда матрица свертки имеет вид:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Умножим (2.2.17) на $\text{vec}(S)$, это выражение свернется в вектор. Затем, преобразовав вектор в обычную матрицу с помощью операции unvec , окончательно получим

$$\begin{aligned} A_{i+1} = A_i - \text{unvec} \left\{ \left(I_n \otimes H_i (A_{i-1} - R P_{i-1}^{-1} A_{i-1}) \right) \cdot \right. \\ \cdot (I_{nm} + P_i^{-1} R \otimes A_i^T P_i^{-1} A_i + E_{nm} (A_i^T P_i^{-1} R \otimes P_i^{-1} A_i) - \\ \left. - P_i^{-1} R \otimes I_m)^{-1} \text{vec}(S) \right\}. \end{aligned} \quad (2.2.18)$$

Выражение (2.2.18) определяет сходящуюся итерационную процедуру, дающую максимально правдоподобные оценки коэффициентов матрицы факторных нагрузок. Однако выбирая различные начальные приближения для матрицы факторных нагрузок в начале итерационной процедуры, в результате получаем практически бесчисленное множество различных, но эквивалентных решений. Все они одинаково описывают корреляции исходных переменных, т.е. формальное перемножение матрицы факторных нагрузок на транспонированную для различных эквивалентных решений будет давать равные матрицы.

Единственное решение может быть найдено лишь при одном условии. Исходная для вычисления факторных нагрузок матрица корреляций является симметричной, таким образом, у нее $\frac{n(n+1)}{2}$ различных элемента. Если учесть, что на главной диагонали корреляционной матрицы стоят единицы, а в случае использования редуцированной корреляционной матрицы на главной диагонали стоят уже оцененные значения характеристик, то для определения $n \cdot m$ неизвестных факторных нагрузок имеется $\frac{n(n-1)}{2}$ исходных величин. Несложно увидеть, что условием единственности факторного решения, по алгебраическим соображениям, является условие равенства числа факторов половине числа переменных, уменьшенного на единицу:

$$m = \frac{n-1}{2}. \quad (2.2.19)$$

Если число переменных – четное, то условие (2.2.19) автоматически невыполнимо. Как правило, $m > \frac{n-1}{2}$, т.е. число неизвестных больше числа уравнений, что ведет, в случае совместности системы уравнений, к бесчисленному множеству решений. Однако и в случае, когда $m < \frac{n-1}{2}$, также возникают проблемы с однозначностью решения. Но даже тогда, когда выполняется условие (2.2.19), составить систему с соответствующим количеством уравнений, имеющих единственное решение, крайне сложно и для большого числа переменных – практически невозможно.

Эти рассуждения построены на алгебраическом подходе. Все же следует отметить, что поиск факторных нагрузок по матрице корреляций тождественен получению собственных векторов симметричной матрицы, а это – решение однородного уравнения, которое имеет бесчисленное множество решений. Все это и есть причина бесчисленности различных эквивалентных максимально правдоподобных факторных решений.

Этот факт – безусловно, относится к недостаткам метода, но большое множество решений позволяет выбрать из них такое, которое бы отвечало дополнительным требованиям. Наиболее распространенное одно из них – это предельная интерпретабельность факторного решения.

С наибольшей наглядностью фактор можно интерпретировать, когда матрица корреляций факторов с переменными, т.е. матрица факторных нагрузок, имеет «простой» вид, а именно, когда элементы в столбце принимают только либо большие, либо малые значения. Причем, поскольку переменные и факторы – нормированы, то «большие» значения – это величины, близкие к единице, а малые значения – к нулю. Разные факторы при этом должны «нагружать» разные переменные таким образом, чтобы, по возможности, «нули» и «единицы» для разных столбцов не совпадали. Такое требование ведет к ортогональности разных столбцов, т.е. скалярное произведение двух разных столбцов равно нулю. Следовательно, в итоге данное требование сводится к тому, чтобы произведение транспонированной матрицы факторных нагрузок на саму себя было диагональной матрицей:

$$A^T A = J, \quad (2.2.20)$$

где J – некоторая диагональная матрица, она в отдельном случае может быть единичной, но поскольку фактор может быть тесно связан с несколькими переменными, и, соответственно, несколько факторных нагрузок могут быть близкими к единице, то матрица J в общем случае – не единичная.

Если настаивать на требовании, чтобы каждый фактор нагружал разные переменные, а факторное решение удовлетворяла бы принципам «простой

структуры», то в среднем на каждый фактор должно приходиться $\frac{n}{m}$ переменных, таким образом, матрица J должна иметь вид:

$$J = \frac{n}{m} I.$$

Для того, чтобы выбрать из множества максимально правдоподобных решений такое, которое удовлетворяло бы ограничению (2.49), необходимо функцию правдоподобия (2.41) дополнить данным ограничением, т.е. составить функцию Лагранжа:

$$g = \ln|P| + \text{tr}(P^{-1}R) + \text{tr}(\Lambda(A^T A - sI)),$$

где Λ – матрица множителей Лагранжа, размером $m \cdot m$,

$s = \frac{n}{m}$ – множитель, соответствующий принципу простой структуры,

остальные обозначения такие же, как и выше.

Теперь найдем производную функции Лагранжа по матрице факторных нагрузок A , затем по матрице множителей Лагранжа Λ . Эти производные после соответствующих преобразований имеют вид:

$$\frac{\partial g}{\partial A} = 2\text{vec}\{P^{-1}A - P^{-1}RP^{-1}A + A\Lambda\};$$

$$\frac{\partial g}{\partial \Lambda} = \text{vec}\{A^T A - sI\}.$$

Эти производные приравняем нулю, и поскольку равенство нулю матрицы, преобразованной в вектор, и самой матрицы имеет одинаковый математический смысл, данные выражения дают систему матричных уравнений:

$$\begin{cases} P^{-1}A - P^{-1}RP^{-1}A + A\Lambda = 0; \\ A^T A - sI = 0. \end{cases} \quad (2.2.21)$$

Преобразуем первое уравнение системы (2.2.21) к выражению, на основании которого можно будет выразить матрицу факторных нагрузок, для этого умножим это уравнение на матрицу истинных корреляций P слева, получим:

$$A - RP^{-1}A + PA\Lambda = 0. \quad (2.2.22)$$

Умножим последнее уравнение (2.2.22) на транспонированную матрицу факторных нагрузок слева, и воспользовавшись вторым уравнением системы (2.2.21), а именно тем, что $A^T A = sI$, получим значение матрицы коэффициентов Лагранжа:

$$\Lambda = (A^T P A)^{-1} (A^T R P^{-1} A - sI). \quad (2.2.23)$$

Далее, подставив матрицу Λ в (2.2.22), и выразив, после очевидных упрощений, матрицу факторных нагрузок, окончательно получим выражение, дающее основание для итерационной процедуры поиска факторного решения:

$$A = R P^{-1} A - P A (A^T P A)^{-1} (A^T R P^{-1} A - sI). \quad (2.2.24)$$

Для начала итерационной процедуры необходима матрица начального приближения факторных нагрузок A_0 , которая может быть получена в соответствии с процедурой (2.2.5). После этого сама итерационная процедура определяется двумя выражениями:

$$\begin{cases} P_i = A_i A_i^T + D_i^2; \\ A_{i+1} = R P_i^{-1} A_i - P_i A_i (A_i^T P_i A_i)^{-1} (A_i^T R P_i^{-1} A_i - sI). \end{cases} \quad (i = 0, 1, \dots) \quad (2.2.25)$$

Причем для получения матрицы P_i нет необходимости вычислять значения характеристистик D_i^2 , поскольку в результате сложения на диагонали матрицы теоретических значений корреляции переменных должны стоять единицы, поэтому после получения произведения $A_i A_i^T$ диагональ этой матрицы заменяется единицами.

Практическое использование процедуры (2.2.25) показало, что процедура сходится достаточно быстро и хорошо, однако для значений факторных нагрузок для четных и нечетных итераций. Т.е. разность между матрицами соседних итераций образует матрицу, элементы которой от итерации к итерации не меняют своих численных значений, после нескольких первоначальных вычислений. Чтобы вывести ход итерации из замкнутого цикла и обеспечить сходимость к одному решению, и вместе с тем удовлетворение условия (2.2.20), процедура (2.2.25) должна быть дополнена еще одним выражением. Его содержание состоит в том, то в качестве входной

величины для следующей итерации берется матрица, лежащая «посередине» между предыдущей матрицей и матрицей, получаемой в ходе уточнения на данной итерации. Таким образом, итерационная процедура, дающее однозначное решение для данного начального значения матрицы факторных нагрузок, имеет следующий вид:

$$\begin{cases} P_i = A_i A_i^T + D_i^2; \\ A_{i+\frac{1}{2}} = R P_i^{-1} A_i - P_i A_i (A_i^T P_i A_i)^{-1} (A_i^T R P_i^{-1} A_i - sI); \\ A_{i+1} = \frac{1}{2} (A_{i+\frac{1}{2}} + A_i); (i = 0, 1, \dots). \end{cases} \quad (2.2.26)$$

Здесь индекс $\frac{1}{2}$ означает промежуточные вычисления. Практическое использование вычислительной процедуры (2.2.26) показало устойчивую сходимость к конечному результату, причем использование в качестве начального приближения различных «хороших» приближений, задаваемых (2.2.25), показало идентичность факторных решений. Единственные отличия состояли в том, что факторы менялись местами: первый становился вторым, третий – первым, а второй – третьим, или их следование осуществлялось в другой комбинации. Причем значения факторных нагрузок существенно не отличались друг от друга. Теоретическое исследование итерационной процедуры говорит о том, что требование выполнения условий простой структуры (2.1.29), приводит к формированию достаточных условий сходимости итерационной процедуры при любом начальном приближении.

Несложно увидеть, что решение, получаемое в ходе процедуры (2.2.26), удовлетворяет условию (2.2.20). А именно, пусть в ходе осуществления итераций получено факторное решение, т.е. матрицы факторных нагрузок соседних итераций не отличаются друг от друга на величину наперед заданной погрешности, и индекс номера итерации можно опустить. Тогда, умножив второе уравнение системы вычислительных тождеств (2.2.26) или выражение (2.2.24) на A^T слева, получим:

$$A^T A = A^T R P^{-1} A - A^T P A (A^T P A)^{-1} (A^T R P^{-1} A - sI) =$$

$$\begin{aligned}
 &= A^T R P^{-1} A - I(A^T R P^{-1} A - sI) = \\
 &= A^T R P^{-1} A - A^T R P^{-1} A + sI = sI.
 \end{aligned}$$

В разделе 2.4 будет показано, что факторное решение (2.2.26) без дополнительных «вращений», которые в факторном анализе традиционно применяются для облегчения трактовки или улучшения интерпретабельности решения, удовлетворяет требованиям простой структуры, и оно четко может быть истолковано. А к ним относятся и такое: каждая переменная может быть отнесена к одному фактору, и число переменных, которые могут быть отнесены к каждому фактору, распределяется равномерно.

2.3. Оценка значений факторов

В анализе экономических процессов важной стороной является изучение поведения факторов во времени. Знание того, как ведут себя скрытые от непосредственного измерения причины взаимодействий между отдельными сторонами явления, делает завершённым ход исследования, позволяя разобраться практически со всеми нюансами поведения экономического объекта и делая прогноз его поведения наиболее достоверным.

Традиционная интерпретация факторного анализа предполагает, что в качестве исходных данных берется не статистика определенных характеристик одного объекта или процесса для различных моментов времени, а наблюдения этих характеристик, произведенные по различным объектам или индивидуумам. И определение значений факторов осуществляется для соответствующих точек исходной информации. Однако исследование экономического процесса предполагает, что исходные данные собираются как фиксация некоторых параметров процесса в определенные моменты времени, либо они получаются как итоговые результаты за определенный промежуток времени (например, годовой валовой доход).

Исследование значений факторов во временном разрезе в традиционном факторном анализе не проводилось. Тем не менее, такое исследование

возможно при определенных предварительных предположениях. Последовательность наблюдений, являющаяся по своей природе выборкой, не должна подпадать под определение «временного ряда», т.е. удовлетворять двум основным условиям:

1) наблюдения в различные моменты времени рассматриваются как случайные величины, являющиеся взаимно независимыми и, в частности, значение, получаемое в некоторый момент времени t_k , ($k = 1, 2, \dots, n$), не должно зависеть от того, какие значения были зарегистрированы до этого момента времени;

2) наблюдения, произведенные в разные моменты времени, образуют стационарную последовательность, т.е. закон распределения вероятности k – го члена выборки (многомерной случайной величины $\mathbf{x}_k = \mathbf{x}(t_k)$) должен оставаться одним и тем же при изменении его номера k , в частности, от t_k не должны зависеть основные числовые характеристики случайной величины $\mathbf{x}_k$ – ее математическое ожидание и дисперсия.

Если эти два требования не выполняются и исходные данные представляют собой временной ряд, то традиционное статистическое исследование такой последовательности производится с помощью специального раздела математической статистики – анализа временных рядов. Впрочем, аппарат факторного анализа может быть использован для решения задач изучения временных рядов, и на это будет указано в третьем разделе. Однако сейчас сделаем предположение о том, что значения показателей экономического процесса в различные моменты времени представляют собой независимые случайные величины, и их совокупность образует стационарную последовательность, т.е. в статистическом плане исходная информация представляет собой некоторую выборку из генеральной совокупности, вероятностные параметры которой не меняются с течением времени.

Различные подходы к оценке значений факторов исследовал Харман [23]. С того момента достижения в этой области не подвергались никакой ревизии,

ни в каком-либо направлении не развивались и сохранили все недостатки, на которые укажем прежде, чем будет предложено кардинальное решение этого вопроса.

Набор из n переменных можно проанализировать либо в терминах только общих факторов, тогда на главной диагонали матрицы корреляций $\mathbf{R}$ стоят единицы, либо в терминах совокупности общих и характерных факторов, и тогда на главной диагонали $\mathbf{R}$ стоят значения общностей. Первый подход приводит к модели компонентного анализа (модели главных компонент): матрица $\mathbf{R}$ – матрица полного ранга, факторное решение дается в терминах n общих факторов и имеет вид

$$\mathbf{Z} = \mathbf{F}\mathbf{A}^T, \quad (2.3.1)$$

где $\mathbf{Z}$ – матрица центрированных и нормированных значений исходных переменных;

$\mathbf{F}$ – матрица значений факторов;

$\mathbf{A}^T$ – транспонированная матрица факторных нагрузок.

Поскольку $\mathbf{A}$ – квадратная невырожденная матрица, то она имеет обратную матрицу. Поэтому искомые значения факторов определяются просто:

$$\mathbf{F} = \mathbf{Z}\mathbf{A}^{-T}. \quad (2.3.2)$$

Это решение является точным и однозначным, и не связано ни с какими «оценками».

При использовании компонентного анализа к факторной модели матрица факторных нагрузок $\mathbf{A}$ является неквадратной, имеющей порядок равный $n \cdot m$. И в этом случае несложно получить значения факторов, умножив (2.3.1) на $\mathbf{A}$ справа:

$$\mathbf{Z}\mathbf{A} = \mathbf{F}\mathbf{A}^T\mathbf{A}, \quad (2.3.3)$$

и решив уравнение (2.3.3) относительно $\mathbf{F}$, найдем значения факторов:

$$\mathbf{F} = \mathbf{Z}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}. \quad (2.3.4)$$

Если матрица A является матрицей собственных значений R , то $A^T A$ – диагональная матрица собственных значений корреляционной матрицы R , и тогда (2.3.4) преобразуется к виду:

$$F = ZAL^{-1}, \quad (2.3.5)$$

где L – диагональная матрица собственных значений R .

Величины значений факторов, полученные по (2.3.5), обладают свойством ортогональности, т.е. матрица $F^T F$ – диагональная:

$$F^T F = L^{-1} A^T Z^T Z A L^{-1} = L^{-1} A^T R A L^{-1} = L^{-1} L L^{-1} = L^{-1}. \quad (2.3.6)$$

Диагональный вид матрицы произведений значений факторов $F^T F$ означает, что процедура (2.3.5) дает ортогональное факторное решение. Однако значения факторов, как и значения преобразованных исходных переменных, должны быть нормированными, т.е. $F^T F$ должна быть равна единичной матрице:

$$F^T F = I, \quad (2.3.7)$$

а матрица L^{-1} в (2.3.6) – обратная матрица собственных значений – не может быть единичной ни для каких исходных данных.

Если, однако, наряду с общими факторами модель включает в себя и характерные факторы, то решение уже нельзя получить столь просто. В этом случае число факторов и общих, и характерных превышает число переменных, и для матрицы факторной структуры не существует обратной. В этом случае традиционным подходом является построение «наилучшего» в смысле наименьших квадратов описания, что подразумевает использование регрессионных методов. Но все же такой подход имеет один существенный недостаток: оценивание с помощью множественной регрессии значений факторов приводит к их коррелированности [26]. С этим утверждением соглашаются вслед и Харман [23, с. 368], и Иберла [19, с. 244]. Тем не менее, можно показать, что полученные с помощью многомерного регрессионного анализа оценки (здесь-таки получаются именно оценки, а не однозначно

определенные по формальной вычислительной процедуре значения) факторов при некоторых условиях являются ортогональными.

В указанных литературных источниках получение значений оценок проведено в такой же степени путано, как и приблизительно. Главный недостаток состоит в следующем: в изложении материала отсутствует ссылка на то, откуда получено главное уравнение, на котором, собственно, и построен этот метод. Не претендуя на новизну, попытаемся раскрыть суть метода, освятив «темные пятна» указанных источников.

Уравнение регрессии, связывающее факторы и переменные можно записать в виде

$$\mathbf{f}_t = \mathbf{z}_t \mathbf{B}, \quad (2.3.8)$$

где $\mathbf{f}_t$ – вектор значений факторов в момент времени t , размером $1 \cdot m$;

$\mathbf{z}_t$ – вектор центрированных и нормированных значений переменных в момент времени t , размером $1 \cdot n$;

$\mathbf{B}$ – матрица неизвестных коэффициентов регрессии, размером $n \cdot m$.

В общем случае линейное уравнение регрессии должно включать и случайную составляющую. Ее наличие не предусматривали «классики», вслед за ними также ею пренебрежем, поскольку природа случайности значений факторов до сих пор не исследовалась.

Умножим (2.3.8) на $\mathbf{z}^T$ слева и возьмем математическое ожидание от всего уравнения:

$$E\{\mathbf{z}^T \mathbf{f}_t\} = E\{\mathbf{z}^T \mathbf{z}_t \mathbf{B}\}.$$

Матрица коэффициентов корреляции является неизменной во времени матрицей, поэтому она, константа, может быть вынесена за знак математического ожидания:

$$E\{\mathbf{z}^T \mathbf{f}_t\} = E\{\mathbf{z}^T \mathbf{z}_t\} \mathbf{B}. \quad (2.3.9)$$

Левая часть (2.3.9) – математическое ожидание произведения переменных на факторы, – представляет собой корреляции между ними, а этими корреляциями является матрица факторных нагрузок $\mathbf{A}$. Математическое

ожидание в правой части (2.3.9) отражает корреляции между переменными, т.е. это математическое ожидание равно матрице корреляций $\mathbf{R}$. Таким образом, выражение (2.3.9) может быть преобразовано к виду:

$$\mathbf{A} = \mathbf{R}\mathbf{B}, \quad (2.3.10)$$

именно вывод этого уравнения отсутствовал у Хармана.

Теперь, поскольку матрица выборочных корреляций с единицами на главной диагонали – полного ранга, то для нее существует обратная, поэтому из (2.3.10) могут быть найдены оценки коэффициентов регрессии

$$\mathbf{B} = \mathbf{R}^{-1}\mathbf{A}. \quad (2.3.11)$$

Оценки значений факторов получаются после подставки коэффициентов регрессии (2.3.11) в уравнение (2.3.8)

$$\mathbf{f}_t = \mathbf{z}_t \mathbf{R}^{-1} \mathbf{A}. \quad (2.3.12)$$

Последнее уравнение связывает значения факторов в момент времени t со значениями переменных в этот же момент. Для всех моментов времени выражение (2.3.12) преобразуется к виду:

$$\mathbf{F} = \mathbf{Z}\mathbf{R}^{-1}\mathbf{A}, \quad (2.3.13)$$

где $\mathbf{F}$ – матрица значений факторов для всех моментов времени;

$\mathbf{Z}$ – матрица центрированных и нормированных значений переменных для всех моментов времени.

Попытаемся проверить ортогональность данной оценки значений факторов, для этого транспонируем матрицу (2.3.13) и затем умножим ее слева на исходную матрицу:

$$\mathbf{F}^T \mathbf{F} = \mathbf{A}^T \mathbf{R}^{-1} \mathbf{Z}^T \mathbf{Z} \mathbf{R}^{-1} \mathbf{A}. \quad (2.3.14)$$

Умножение на множитель $\frac{1}{N-1}$ выражения (2.3.14) делает произведение $\mathbf{Z}^T \mathbf{Z}$ матрицей выборочных корреляций $\mathbf{R}$, поскольку произведение обратной матрицы на исходную есть матрица единичная, то (2.3.14) преобразуется к виду:

$$\mathbf{F}^T \mathbf{F} = \mathbf{A}^T \mathbf{R}^{-1} \mathbf{A}. \quad (2.3.15)$$

Если для оценки матрицы факторных нагрузок $\mathbf{A}$ используется метод главных факторов – наиболее распространенный до сего момента времени метод, – то $\mathbf{A}$ состоит из собственных векторов матрицы выборочных корреляций. И поскольку обратная матрица в (2.3.15) имеет те же собственные векторы, что и исходная, то (2.3.15) преобразуется к виду:

$$\mathbf{F}^T \mathbf{F} = \mathbf{L}^{-1}, \quad (2.3.16)$$

где $\mathbf{L}$ – матрица собственных значений выборочной корреляционной матрицы.

Таким образом, этот метод дает такой же результат, как и компонентный метод, и корреляции между значениями факторов, вычисленных с использованием метода главных факторов, аналогичны (2.3.6); и, соответственно, замечание о коррелированности факторов при применении методов многомерного регрессионного анализа – не совсем корректно. Тем не менее, применение других методов оценки матрицы факторных нагрузок, кроме метода главных факторов, когда столбцы факторных нагрузок не есть собственные векторы корреляционной матрицы, ведет к нарушению ортогональности факторной структуры. Следовательно, не само использование многомерной регрессии ведет к коррелированности факторов, а к этому приводят результаты использования соответствующего метода оценки факторных нагрузок.

Харман [23] и Иберла [19] подробно рассмотрели вопросы проверки точности оценки значений факторов. Этот момент является достаточно важным, поэтому рассмотрим теорию проблемы в изложении «классиков».

Прежде всего, они исходили из того, что матрица факторных нагрузок $\mathbf{A}$ является ортогональной т.е. ее столбцы – собственные векторы матрицы корреляций. Как теперь известно, это ведет к ортогональности значений факторов.

Если известна оценка матрицы коэффициентов регрессии, то подставляя в (2.3.8) значение переменных для некоторого момента времени, получим оценку значения фактора для данного момента. Ставится задача: насколько

точно такая оценка? Мерой качества такой оценки могут служить коэффициент множественной корреляции и коэффициент множественной детерминации [19, с. 251].

При парной линейной связи между лишь двумя переменными коэффициент корреляции характеризует тесноту связи между двумя переменными. Причем чем ближе его значение по абсолютной величине к единице, тем теснее связь между переменными. Соответствующим показателем во множественной регрессии является коэффициент множественной корреляции, или совокупный коэффициент корреляции. Он показывает тесноту связи зависимой переменной, или целевой функции, от переменных, входящих в правую часть уравнения регрессии.

Проверка точности оценки значения фактора осуществляется на основании корреляции между истинным, но неизвестным значением фактора и его оценкой, но поскольку оценка вычисляется по уравнению регрессии, то критерий качества оценки получается на основании коэффициента множественной корреляции между истинным значением фактора и значениями переменных, на основании которых получена оценка.

Для фактора F_j коэффициент множественной корреляции имеет вид [23, с. 373]

$$R_j = \sqrt{a_{j,1}b_{j,1} + a_{j,2}b_{j,2} + \dots + a_{j,n}b_{j,n}}, \quad (2.3.17)$$

где $a_{j,i}$ – элемент матрицы факторных нагрузок $\mathbf{A}$,

$b_{j,i}$ – элемент матрицы коэффициентов множественной регрессии $\mathbf{B}$.

Квадрат коэффициента множественной корреляции называется коэффициентом множественной детерминации. В соответствии с этим определением для фактора F_j коэффициент множественной детерминации имеет вид

$$R_j^2 = a_{j,1}b_{j,1} + a_{j,2}b_{j,2} + \dots + a_{j,n}b_{j,n}. \quad (2.3.18)$$

Если учесть (2.3.11), т.е. принять во внимание формулу оценки коэффициентов регрессии, то (2.3.18) можно представить в следующем виде:

$$R_j^2 = \mathbf{a}_j^T \mathbf{R}^{-1} \mathbf{a}_j, \quad (2.3.19)$$

где $\mathbf{a}_j$ – j -й столбец матрицы факторных нагрузок.

Коэффициент множественной детерминации состоит из нескольких слагаемых, каждое из которых вносит свой вклад в R_j^2 . Каждое из этих слагаемых в свою очередь является произведением коэффициента регрессии и коэффициентом корреляции между j -м фактором и i -й переменной. Если какое-либо слагаемое в (2.3.18) значительно по величине, то соответствующая переменная вносит большой вклад в определение фактора F_j . По коэффициенту множественной детерминации можно определить, какая часть дисперсии фактора обусловлена переменными, а какая часть обусловлена неучтенными переменными.

Эти же соображения можно отнести и к влиянию фактора на переменные: большое значение коэффициента детерминации говорит о значимости данного фактора в формировании поведения переменных. И *contra versa* – небольшое значение этого коэффициента свидетельствует о незначимости данного фактора, а в решении проблемы точности оценки значений факторов – о недостаточно хорошем качестве оценки значения фактора.

Коэффициент множественной детерминации, как и коэффициент множественной корреляции, принимает значение не большее, чем единица. Поэтому близость к единице будет свидетельствовать о значимости данного фактора и достаточно хорошей оценке его значения. К тому же коэффициент множественной корреляции позволяет проверить гипотезу о значимости этого коэффициента с заданным уровнем значимости (вероятностью ошибки), что позволяет объективировать процесс проверки меры близости оценки фактора к его истинному значению.

Проверка осуществляется с помощью статистической таблицы «Критические значения коэффициента множественной корреляции». Для данного уровня значимости, заданного числа переменных и наблюдений вычисленный коэффициент множественной корреляции сравнивается с его

табличным критическим значением. Если вычисленный коэффициент больше табличного, то его истинное значение отлично от нуля.

Табличное значение задается на основании распределения Фишера по формуле:

$$\rho_{k_1, k_2}^{(\alpha)} = \sqrt{\frac{1}{1 + \frac{1}{k_1 \cdot F(k_1, k_2)}}}, \quad (2.3.20)$$

где $\rho_{k_1, k_2}^{(\alpha)}$ – критическое значение множественного коэффициента корреляции с α – уровнем значимости и k_1, k_2 степенями свободы;

k_1 – число переменных;

$k_2 = N - k_1 - 1$, где N – число наблюдений (объем выборки);

$F(k_1, k_2)$ – распределение Фишера с k_1, k_2 степенями свободы и с тем же, что и у критического значения коэффициента корреляции, уровнем значимости.

Если гипотеза о значимости коэффициента множественной корреляции отклоняется, то соответствующий фактор интерпретировать в принципе нельзя, поскольку и вклад переменных в информацию о нем незначителен. Такое положение несложно исправить, уменьшив число факторов на единицу. Правда, проверка на значимость с помощью таблицы критических значений коэффициента множественной корреляции не совсем корректна, так она составлена в предположении независимости переменных, с которыми имеется корреляция некоторого признака [19, с. 252]. Но коррелированность «независимых» переменных ведет к завышению оценки коэффициента множественной корреляции. Это следует из известного в теории вероятностей утверждения о вероятности суммы произвольных, в том числе и зависимых событий, согласно которому вероятность суммы произвольных событий равна сумме вероятностей этих событий, уменьшенной на вероятность произведения этих событий:

$$P(A + B) = P(A) + P(B) - P(AB), \quad (2.3.21)$$

где $P(\cdot)$ – вероятность события;

$A + B$ – сумма зависимых событий;

AB – произведение событий, т.е. вероятность того, что они одновременно произойдут, а для независимых событий эта вероятность равна произведению вероятностей этих событий:

$$P(AB) = P(A) \cdot P(B), \quad (2.3.22)$$

но события зависимы, и для них сумма вероятностей все одно уменьшается на вероятность их совместного наступления.

Случайная величина ξ тесно связана с вероятностью события состоящего в том, что ξ примет то или иное значение. И поскольку оценка той или иной случайной величины несет в себе информацию об истинном, но неизвестном ее значении, а информация тесно зависит от вероятности, то отрицательная часть (2.3.21) ведет к ее снижению. Поясним – почему. Левая часть (2.3.21) – вероятность суммы событий – представляет собой вероятность того, что фактор связан с несколькими переменными, которые, в общем-то, взаимозависимы. Т.е. вероятность, стоящая в левой части (2.3.21), связана со значением фактора. В правой части (2.3.21) – сумме вероятностей событий – каждое слагаемое несет вероятность того, что отдельная переменная формируется фактором, и этот момент отмечен величиной слагаемого. Но общая сумма вероятностей должна быть скорректирована на вероятность их произведения. Так как переменные в определенный момент t проявляются практически все, то такая коррекция должна иметь значительный вес. Поэтому табличные критические значения коэффициента множественной корреляции составлены для независимых переменных, в том числе и потому, что коррекция на вероятность их совместного наступления в практическом плане зависит от специфики самих переменных и в общем случае учтена быть не может.

Завышение оценки значимости фактора может вести лишь к одному результату: в факторном анализе будут появляться факторы, значимость которых подтверждается критерием, но в реальности значимыми они не являются, т.е. соответствующим оценкам значений фактора верить не приходится. Поэтому решение этой проблемы должно лежать в плоскости

минимизации числа факторов, минимизации – не как оптимизационной задачи, а выбора минимально необходимого для адекватного анализа числа факторов.

Небольшой промежуточный вывод: оценка значений факторов, основанная на методе множественной регрессии и осуществляемая в соответствии с выражением (2.3.13), дает ортогональные величины факторов, но они не являются нормированными, т.е. сумма квадратов значений отдельного фактора не равно единице. Это существенный недостаток, поскольку в факторной модели нормированные значения переменных выражаются через значения факторов, и если последние – не нормированы, то и сами переменные нормированными быть не могут. С формальной точки зрения это противоречие преодолено быть не может, поэтому в последнее время появился подход к модели факторного анализа, позволяющий «зрительно» обойти эту проблему [27]. В факторную модель включают центрированные, но не нормированные величины, поэтому в самой модели присутствуют средние значения переменных, отклонения от которых задают факторы:

$$\mathbf{z}_t = \bar{\mathbf{z}} + \mathbf{f}_t \mathbf{A}^T + \mathbf{u}_t, \quad (2.3.23)$$

где $\mathbf{z}_t$ – в этом контексте, – вектор центрированных переменных в момент времени t , т.е. у значений переменных вычтены средние значения;

$\bar{\mathbf{z}}$ – вектор средних значений переменных;

$\mathbf{f}_t$ – вектор значений факторов в момент времени t , в этом контексте значения факторов считаются центрированными, но не нормированными;

$\mathbf{A}$ – матрица факторных нагрузок;

$\mathbf{u}_t$ – вектор случайных отклонений в момент времени t , дисперсии которых составляют характеристики для переменных.

Идеальным в факторном анализе выглядит ситуация, когда модель соотношения переменных и факторов дается в натуральных единицах, тогда проблема с центрированием и нормированием выглядела бы надуманной. Но есть один вопрос, на который очень сложно ответить: в каких единицах должен измеряться фактор так, чтобы через него выражались переменные разной

размерности? К тому же одна переменная выражается через совокупность факторов. Эта практически не разрешимая проблема как раз и снимается с помощью центрирования и нормирования. Причем одно лишь центрирование проблемы не снимает, поскольку, если из некоторой величины вычесть среднее значение, имеющее ту же единицу измерения, то от этого размерность центрированной переменной ни исчезнет. От размерности освобождаются лишь при делении центрированной величины на среднееквадратическое отклонение.

Поэтому модель (2.3.23) неправомерна. Значение фактора должно быть безразмерным и также центрированным, как центрированы исходные переменные. Причем вернуться к натуральному измерению переменных не составляет труда, если известны их средние величины и среднееквадратические отклонения, но, опять же, значения факторов должны быть и центрированными (имеющими нулевое математическое ожидание) и нормированными (деленными на собственные среднееквадратические отклонения).

Поставим задачу получить оценки значений факторов для всех моментов времени, математические ожидания которых были бы равны нулю, а дисперсии – единице. Т.е. получить так называемые «стандартные» значения факторов.

Напомним основную модель факторного анализа для всех моментов:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}^T + \mathbf{U}\mathbf{D}, \quad (2.3.24)$$

где $\mathbf{Z}$ – матрица размером $N \cdot n$ наблюдений переменных, ее значения центрированы и нормированы;

$\mathbf{F}$ – матрица размером $N \cdot t$ значений факторов, именно ее значения предстоит оценить;

$\mathbf{A}$ – матрица размером $n \cdot t$ факторных нагрузок, будем предполагать, что оценка ее элементов уже получена с помощью метода максимального правдоподобия, изложенного в п.2.2;

$\mathbf{U}$ – матрица размером $N \cdot n$ значений характерных факторов.

$\mathbf{D}$ – матрица размером $n \cdot n$ нагрузок при характерных факторах

Выше рассматривалась модель (1.1.39), которая ничем не отличается от (2.3.24), однако внесем в нее некоторые коррективы. Поскольку поставлена задача получения оценок значения лишь общих факторов, то второе слагаемое – влияние на поведение переменной характерных факторов, – можно рассматривать как некоторые случайные отклонения от основной зависимости, задаваемой общими факторами. В этой связи перепишем модель (2.3.24) в виде:

$$\mathbf{Z} = \mathbf{F}\mathbf{A}^T + \mathbf{V}, \quad (2.3.25)$$

где $\mathbf{V} = \mathbf{U}\mathbf{D}$ – матрица случайных отклонений, математическое ожидание которой, как несложно увидеть, равно нулю, а дисперсия – $\mathbf{D}^2$, т.е. значения характерностей.

Очевидным представляется стремление получить такие оценки значений факторов, которые минимизировали бы ошибки в описании переменных посредством факторной модели. При этом ошибками будем считать отклонения реальных значений переменных от теоретической зависимости, определяемой значениями факторов:

$$\hat{\mathbf{V}} = \mathbf{Z} - \mathbf{F}\mathbf{A}^T, \quad (2.3.26)$$

где $\hat{\mathbf{V}}$ – ошибки в объяснении переменных посредством факторов. Эти ошибки могут рассматриваться как оценки неизвестных значений характерностей в общей факторной модели.

В теории статистического оценивания широко известен метод наименьших квадратов оценивания параметров регрессии, который позволяет минимизировать сумму квадратов ошибок, т.е. отклонений реальных значений результирующего признака от «идеальной» регрессионной зависимости этого признака, определяемой некоторой группой переменных. В нашем случае выражение (2.3.25) можно считать уравнением регрессии, в котором значения факторов $\mathbf{F}$ играют роль неизвестных параметров, оценки которых подлежат определению, матрица факторных $\mathbf{A}$ нагрузок может быть интерпретирована как матрица значений независимых переменных, $\mathbf{Z}$ – матрица значений зависимой переменной.

Необходимым условием получения оценок является условие достаточности исходных данных, при котором числа уравнений должно хватить для определения неизвестных. В нашем случае размерность только одной матрицы $\mathbf{Z}$, элементы которой составляют исходные данные, равна $N \cdot n$, а число неизвестных – элементов матрицы значений факторов $\mathbf{F}$ составляет $N \cdot t$, где t – число факторов – значительно меньше числа переменных n . Даже если учесть, что часть информации «израсходована» на получение оценки факторных нагрузок, число которых равно $n \cdot t$, исходных данных все равно достаточно для получения оценок значений факторов.

Основным соотношением для решения вопроса о достаточности первоначальной информации является следующее: наблюдений за переменными больше числа значений факторов плюс число факторных нагрузок:

$$N \cdot n > N \cdot t + n \cdot t. \quad (2.3.27)$$

Несложно показать, что даже при максимально возможном числе факторов, составляющем половину от числа переменных, условие (2.3.27) выполняется. Пусть $t = \frac{n}{2}$, тогда $N \cdot n > N \cdot \frac{n}{2} + n \cdot \frac{n}{2}$; $N > N \cdot \frac{1}{2} + n \cdot \frac{1}{2}$; $\frac{N}{2} > \frac{n}{2}$; $N > n$; а условие, когда число наблюдений больше числа исследуемых величин, является основополагающим требованием статистических исследований, и выполняется всегда. При этом следует отметить, что в практических исследованиях число интерпретабельных факторов, которые и составляют основу факторной модели конкретного явления, не бывает больше четверти числа переменных.

Таким образом, в качестве целевой функции при получении оценок значений факторов можно выбрать сумму квадратов ошибок (2.3.26), для матрицы (не для вектора или матрицы-строки) она имеет вид:

$$tr\{(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)^T(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)\}, \quad (2.3.28)$$

где tr – след матрицы, т.е. сумма диагональных элементов матрицы.

Однако необходимо найти не просто значения $\mathbf{F}$, которые минимизировали бы (2.3.28), но эти значения должны быть нормированными, т.е. обладать свойством (2.3.7): $\mathbf{F}^T \mathbf{F} = \mathbf{I}$. Для поиска минимума целевой функции при условии наложенных ограничений (2.3.7) составим функцию Лагранжа:

$$g = \text{tr}\{(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)^T(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)\} + \text{tr}\{\mathbf{\Lambda}(\mathbf{F}^T \mathbf{F} - \mathbf{I})\}, \quad (2.3.29)$$

где $\mathbf{\Lambda}$ – матрица множителей Лагранжа.

Найдем экстремумы функции (2.3.29), для этого вычислим частные производные g по $\mathbf{F}$ и по $\mathbf{\Lambda}$:

$$\begin{aligned} \frac{\partial g}{\partial \mathbf{F}} &= \frac{\partial}{\partial \mathbf{F}} (\text{tr}\{(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)^T(\mathbf{Z} - \mathbf{F}\mathbf{A}^T)\} + \text{tr}\{\mathbf{\Lambda}(\mathbf{F}^T \mathbf{F} - \mathbf{I})\} + \text{tr}\{\mathbf{\Lambda}(\mathbf{F}^T \mathbf{F} - \mathbf{I})\}) = \\ &= \frac{\partial}{\partial \mathbf{F}} (\text{tr}\{\mathbf{Z}^T \mathbf{Z} - \mathbf{Z}^T \mathbf{F}\mathbf{A}^T - \mathbf{A}\mathbf{F}^T \mathbf{Z} + \mathbf{A}\mathbf{F}^T \mathbf{F}\mathbf{A}^T\} + \text{tr}\{\mathbf{\Lambda}\mathbf{F}^T \mathbf{F} - \mathbf{\Lambda}\}) = \\ &= \{\mathbf{E}(\mathbf{A}^T \otimes \mathbf{F}\mathbf{A}^T) + (\mathbf{F}\mathbf{A}^T \otimes \mathbf{A}^T) - \mathbf{E}(\mathbf{A}^T \otimes \mathbf{Z}) - (\mathbf{Z} \otimes \mathbf{A}^T) + \mathbf{E}(\mathbf{\Lambda}^T \otimes \mathbf{F}) + \\ &\quad + (\mathbf{F}\mathbf{\Lambda}^T \otimes \mathbf{I})\} \text{vec}\{\mathbf{I}\} = \text{vec}\{2\mathbf{F}\mathbf{A}^T \mathbf{A} - 2\mathbf{Z}\mathbf{A} + 2\mathbf{F}\mathbf{\Lambda}\}; \\ \frac{\partial g}{\partial \mathbf{\Lambda}} &= \text{vec}\{\mathbf{F}^T \mathbf{F} - \mathbf{I}\}. \end{aligned}$$

Приравняем нулю найденные частные производные, и после очевидных преобразований получим систему матричных уравнений:

$$\begin{cases} -\mathbf{Z}\mathbf{A} + \mathbf{F}(\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda}) = \mathbf{0} \\ \mathbf{F}^T \mathbf{F} = \mathbf{I} \end{cases}. \quad (2.3.30)$$

Из первого уравнения системы (2.3.30) выразим матрицу $\mathbf{F}$, и найденное значение подставим во второе уравнение системы:

$$\begin{aligned} \mathbf{F} &= \mathbf{Z}\mathbf{A}(\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda})^{-1}, \\ \mathbf{F}^T &= (\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda})^{-1} \mathbf{A}^T \mathbf{Z}^T, \\ (\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda})^{-1} \mathbf{A}^T \mathbf{Z}^T \mathbf{Z}\mathbf{A}(\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda})^{-1} &= \mathbf{I}. \end{aligned} \quad (2.3.31)$$

Выразим матрицу, стоящую в середине левой части, умножив слева и справа на обратную матрицу, содержащую множители Лагранжа, получим

$$\mathbf{A}^T \mathbf{Z}^T \mathbf{Z}\mathbf{A} = (\mathbf{A}^T \mathbf{A} + \mathbf{\Lambda})^2.$$

Извлечение квадратного корня дает следующее выражение:

$$(\mathbf{A}^T \mathbf{Z}^T \mathbf{Z}\mathbf{A})^{\frac{1}{2}} = \mathbf{A}^T \mathbf{A} + \mathbf{\Lambda},$$

откуда получаем матрицу множителей Лагранжа

$$\Lambda = (A^T Z^T Z A)^{\frac{1}{2}} - A^T A. \quad (2.3.32)$$

Подставляя найденную матрицу (2.3.32) в выражение (2.3.31), получим оценку значений факторов

$$\hat{F} = Z A (A^T Z^T Z A)^{-\frac{1}{2}}, \quad (2.3.33)$$

которая должна быть ортогональной, т.е. значения факторов – нормированы, и, соответственно, ее математическое ожидание должно быть равно нулю. Проверим эти свойства.

Сначала установим, чему равно математическое ожидание оценки. Предварительно отметим, что поскольку исходные переменные центрированы и нормированы, то математическое ожидание матрицы таких значений равно нулю:

$$\mathcal{M}\{Z\} = 0,$$

где $\mathcal{M}\{\cdot\}$ – символ математического ожидания. При этом следует отметить, что при выполнении условий центрирования и нормирования, произведение матриц $Z^T Z$ равно корреляционной матрице исходных переменных R . Это замечание должно быть сделано, для того, чтобы обойти проблему, связанную с определением математического ожидания обратной случайной величины. Если математическое ожидание некоторой случайной величины равно, скажем a , то есть $\mathcal{M}\{\xi\} = a$, то математическое ожидание обратного значения случайной величины в общем случае не равно обратному значению математического ожидания: $\mathcal{M}\left\{\frac{1}{\xi}\right\} \neq \frac{1}{a}$. И поскольку матрица Z находится под знаком обратной величины в выражении (2.3.33), то такая проблема имеет место. Решение этой проблемы лежит за пределами данного исследования, поэтому положим то, что с чего начинается факторный анализ – с вычисления корреляционной матрицы. И так как в (2.3.33) под знаком обратной величины находится не просто матрица Z , а ее момент второго порядка $Z^T Z$, то с точностью до скалярных множителей далее будем полагать, что:

$$\mathbf{Z}^T \mathbf{Z} = \mathbf{R},$$

Что касается множителей, то имеется в виду множитель, равный $\frac{1}{N-1}$. На этот множитель можно умножить обе части (2.3.33), от чего это тождество не потеряет своей справедливости. Причем $\mathbf{R}$ – выборочная корреляционная матрица, которая по отношению к оператору взятия математического ожидания может считаться постоянной. Данное обстоятельство должно быть доказано, или доказательно пояснено, причем это возможно, но находится за пределами данного исследования. Поэтому уравнение (2.3.33) можно переписать в виде

$$\hat{\mathbf{F}} = \mathbf{Z}\mathbf{A}(\mathbf{A}^T \mathbf{R}\mathbf{A})^{-\frac{1}{2}}. \quad (2.3.34)$$

А теперь можно найти математическое ожидание оценки факторных нагрузок, учитывая, что постоянную величину можно вынести за знак математического ожидания:

$$\mathcal{M}\{\hat{\mathbf{F}}\} = \mathcal{M}\left\{\mathbf{Z}\mathbf{A}(\mathbf{A}^T \mathbf{R}\mathbf{A})^{-\frac{1}{2}}\right\} = \mathcal{M}\{\mathbf{Z}\}\mathbf{A}(\mathbf{A}^T \mathbf{R}\mathbf{A})^{-\frac{1}{2}} = \mathbf{0} \cdot \mathbf{A}(\mathbf{A}^T \mathbf{R}\mathbf{A})^{-\frac{1}{2}} = \mathbf{0}.$$

Таким образом, оно равно нулю, что и требовалось доказать. А так как истинные, но неизвестные значения центрированных и нормированных факторов равны нулю, то доказано утверждение, что данные оценки являются несмещенными.

Покажем теперь, что оценка (2.88) является ортонормированной, для этого найдем произведение $\hat{\mathbf{F}}^T \hat{\mathbf{F}}$:

$$\hat{\mathbf{F}}^T \hat{\mathbf{F}} = (\mathbf{A}^T \mathbf{Z}^T \mathbf{Z} \mathbf{A})^{-\frac{1}{2}} \mathbf{A}^T \mathbf{Z}^T \mathbf{Z} \mathbf{A} (\mathbf{A}^T \mathbf{Z}^T \mathbf{Z} \mathbf{A})^{-\frac{1}{2}} = (\mathbf{A}^T \mathbf{Z}^T \mathbf{Z} \mathbf{A})^{-\frac{1}{2}} (\mathbf{A}^T \mathbf{Z}^T \mathbf{Z} \mathbf{A})^{\frac{1}{2}} = \mathbf{I}.$$

Поскольку это произведение равно единичной матрице, то оценка значений факторов является ортогональной, и это – ставилось как основное требование при получении этой оценки.

Следовательно, оценка значений факторов (2.3.33) является несмещенной и ортонормированной.

Ортонормированные оценки являются наилучшими оценками среди всех прочих оценок, поскольку ортогональность предполагает взаимную независимость факторов, что отражает основное исходное положение

факторного анализа – независимость причин, вызывающих взаимодействия между переменными, и если факторы коррелируют друг с другом, то это означает, что над ними стоит «надфактор», вызывающий такое явление и подлежащий выявлению. Нормированность делает возможным как получение на основании значений факторов таких же нормированных значений переменных, так и использование классической модели факторного анализа, а не различных модификаций типа (2.78), не имеющих физического смысла, строго говоря.

2.4. Исследование формирования остатков на депозитных счетах

В одном из коммерческих банков г. Севастополя относительно недавно был поставлен уникальный эксперимент по привлечению денежных средств населения на депозитные счета. В конкурентной борьбе за клиента этот банк, назовем его – КБ, принял уникальную программу по резкому увеличению процентных ставок по депозитным договорам, сначала по наиболее популярным срокам – до года, а затем и по договорам со сроками до двух и более лет. Имеются данные за пятнадцать месяцев по суммам остатков на депозитных счетах на начало соответствующего месяца коммерческого банка, приведенные в таблице 2.4.1.

Таблица 2.4.1 – Остатки на депозитных счетах и процентные ставки КБ

Месяц	Депозиты (тыс. грн.)				Ставка процента (%)			
	Нац. валюта		Ин. валюта		Нац. валюта		Ин. валюта	
	2630	2635	2630	2635	2630	2635	2630	2635
1	1424	3	2165	2124	16,69	13,13	24,02	21,00
2	1518	3	2164	2145	16,69	13,13	23,52	21,00
3	1627	3	2194	1949	16,83	12,98	23,38	21,00
4	1668	3	2292	1960	16,52	12,99	23,11	21,00

Продолжение таблицы 2.4.1

5	1564	36	2292	978	16,24	15,72	22,84	24,38
6	1307	1585	2197	2216	15,71	17,60	22,29	24,26
7	1350	2390	2232	2463	15,68	18,49	22,00	24,30
8	1502	3264	2550	2924	15,39	18,84	21,59	24,29
9	1523	4164	2644	3198	14,94	19,16	20,86	24,15
10	1670	5210	2165	3400	14,69	18,93	20,11	23,94
11	1772	7129	2019	4686	14,71	18,86	19,98	23,42
12	1824	7879	1972	4784	14,86	19,00	19,94	23,32
13	1857	8744	2010	5099	14,67	18,99	19,62	23,14
14	2097	10011	2266	5339	14,39	19,05	19,47	22,91
15	2396	10894	2247	5422	14,49	19,13	19,12	22,76

Исходные данные имеют следующие очевидные особенности. От начального момента исследования процентные ставки по депозитам физических лиц в национальной валюте до года (счет 2630) только снижались, а процентные ставки по депозитам свыше года (счет 2635) в течение первых четырех месяцев не изменялись, а затем резко выросли и, через некоторый промежуток времени их уровень оказался выше, чем на счетах 2630.

Небольшое, но непрерывное уменьшение ставок на счетах 2630 не привело к уменьшению или даже замедлению процесса накопления средств на этих счетах. Резкое увеличение ставки на 2635 счете не сразу сказалось на росте денежных средств, однако, последующее плавное снижение ставки опять же не привело к сокращению притока средств, а напротив, наблюдался их постоянный прирост. Очевидно, что за всеми этими процессами стоят непосредственно не измеряемые факторы. Проведем факторный анализ процесса формирования денежных средств на депозитных счетах в зависимости от уровня процентной ставки и ее изменения.

Приведем те исходные данные, которые были использованы в факторном анализе динамики депозитных остатков, а именно: исходные величины были

отцентрированы и пронормированы. Т.е. из каждого наблюдения было вычтено среднее значение данной переменной, а затем эта разность была разделена на среднеквадратическое отклонение переменной. Рассматривались следующие показатели:

X_1 – депозиты физических лиц в национальной валюте на первое число очередного месяца по счету 2630;

X_2 – изменение X_1 , т.е. разность текущего и предшествующего значений величины депозитов;

X_3 – депозиты физических лиц в национальной валюте на первое число очередного месяца по счету 2635;

X_4 – изменение X_3 ;

X_5 – процентная ставка счета 2630;

X_6 – изменение X_5 ;

X_7 – процентная ставка счета 2635;

X_8 – изменение X_7 .

Таблица 2.4.2 – Исходные данные для факторного анализа депозитов в национальной валюте

Месяц	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	-0,597	0,181	-1,107	-1,305	-1,419	-0,689	-1,672	-0,133
2	-0,221	0,291	-1,107	-1,305	1,331	0,964	-1,672	-0,133
3	-0,080	-0,209	-1,107	-1,305	1,160	0,367	-1,672	-0,133
4	-0,438	-1,276	-1,099	-1,249	0,990	0,367	0,977	3,435
5	-1,325	-2,403	-0,707	1,293	0,642	-0,918	0,883	-0,259
6	-1,177	-0,195	-0,503	-0,045	0,459	0,276	0,914	-0,090
7	-0,652	0,608	-0,282	0,161	0,200	-0,276	0,906	-0,143
8	-0,580	-0,356	-0,055	0,205	-0,261	-1,745	0,797	-0,280
9	-0,073	0,571	0,210	0,450	-0,735	-1,837	0,632	-0,354
10	0,279	0,240	0,696	1,914	-0,817	1,010	0,224	-0,681
11	0,459	-0,128	0,885	-0,047	-0,842	1,424	0,146	-0,238

Продолжение таблицы 2.4.2

12	0,572	-0,268	1,104	0,146	-1,045	0,138	0,005	-0,323
13	1,400	1,255	1,425	0,820	-1,139	0,918	-0,175	-0,375
14	2,432	1,690	1,648	0,176	-1,361	0	-0,293	-0,291

В таблице указано четырнадцать значений месяцев. Для переменной изменения показателя необходимо наблюдений на единицу больше, поэтому для вычисления первых величин показателей изменения использовались данные предшествующего месяца, не приведенного в таблице.

Между переменными оказались следующие парные коэффициенты корреляции (таблица 2.4.3).

Таблица 2.4.3 – Матрица корреляций между переменными депозитов в национальной валюте

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1	0,716	0,803	0,144	-0,700	0,348	-0,238	-0,182
X_2	0,716	1	0,554	0,013	-0,480	0,180	-0,311	-0,383
X_3	0,803	0,554	1	0,594	-0,972	0,200	0,215	-0,398
X_4	0,144	0,013	0,594	1	-0,668	-0,082	0,564	-0,465
X_5	-0,700	-0,480	-0,972	-0,668	1	-0,064	-0,368	0,377
X_6	0,348	0,180	0,200	-0,082	-0,064	1	-0,292	0,089
X_7	-0,238	-0,311	0,215	-0,368	-0,368	-0,292	1	0,238
X_8	-0,182	-0,383	-0,398	0,377	0,377	0,089	0,238	1

Число факторов было выбрано, равное трем, исходя из рекомендаций по формированию «простой структуры» факторного решения.

Факторный анализ, проведенный для трех факторов, дал максимально правдоподобные оценки факторных нагрузок, приведенные в таблице 2.4.4

Таблица 2.4.4 – Факторные нагрузки переменных депозитов в национальной валюте

	F_1	F_2	F_3
X_1	0,517	-0,298	-0,037
X_2	0,321	-0,186	-0,300
X_3	0,183	-0,489	0,023
X_4	-0,366	-0,456	0,017
X_5	-0,083	0,517	-0,085
X_6	0,499	-0,002	0,074
X_7	-0,297	-0,314	0,601
X_8	0,344	0,253	0,731

Теперь проведем интерпретацию факторного решения. Поскольку решение получено с учетом требований «простой структуры», то каждая переменная должна быть отнесена лишь к одному из факторов. Поэтому в каждой строке матрицы факторных нагрузок выберем наибольшее значение. В таблице 2.4.5 эти значения выделены фоновым цветом.

Таблица 2.4.5 – Отнесение каждой переменной к отдельному фактору

	F_1	F_2	F_3
X_1	0,517	-0,298	-0,037
X_2	0,321	-0,186	-0,300
X_3	0,183	-0,489	0,023
X_4	-0,366	-0,456	0,017
X_5	-0,083	0,517	-0,085
X_6	0,499	-0,002	0,074
X_7	-0,297	-0,314	0,601
X_8	0,344	0,253	0,731

Как видно из таблицы 2.4.5, также выполнено требование равномерности распределения переменных по отношению к факторам.

Теперь, первый фактор определяется величиной депозитов на счете 2630, изменением этой величины, а также изменением процентной ставки этого счета. Таким образом, первый фактор наиболее связан с депозитами до одного года. Заметим, что среди остальных факторных нагрузок самое большое значение, большее даже, чем значение корреляции между определяющей переменной X_2 и фактором F_1 , имеет коэффициент $a_{4,1}$ – нагрузка первого фактора на изменение остатков на депозитах свыше года, причем нагрузка – отрицательная. Т.е. в интерпретации первого фактора этот момент должен быть использован. Также достаточно большое значение имеет коэффициент $a_{8,1}$ – показатель связи между первым фактором и изменением процентной ставки по счету 2635.

В совокупности первый фактор очень близок к общему фактору. Поэтому он может быть интерпретирован, как фактор быстроты получения доходов от депозитных вкладов. Причем, чем меньше срок до выплаты вклада, тем лучше с точки зрения этого фактора. Первый фактор угнетает рост остатков на депозитах с большим сроком, и способствует положительному изменению (знак «плюс» у коэффициента $a_{8,1}$), т.е. росту процентных ставок на более срочных депозитах.

Второй фактор определяется величиной остатков на долгосрочных депозитных счетах, их изменением, причем на эти переменные этот фактор оказывает негативное влияние; а также – величиной процентных ставок на депозитах до года. Это так называемые определяющие переменные. Среди переменных, имеющих большую факторную нагрузку, – процентная ставка по долгосрочным депозитам, также со знаком минус.

Следует отметить, что знак у коэффициента факторной нагрузки имеет значение, поскольку максимально правдоподобные оценки находились при условии ортогональности факторных нагрузок, как требования «простой

структуры». Казалось бы, изменение знака на противоположный при интерпретации сделает этот процесс более простым, однако факторные нагрузки потеряют ортогональность и сами факторы тогда не могут быть рассматриваемы, как независимые причины. Поменять знаки у коэффициентов факторных нагрузок можно только целиком у всей матрицы, а не у отдельного ее столбца.

С учетом знаков при переменных, имеющих наибольшие нагрузки, а среди них – показатели, характеризующие оба депозитных счета; второй фактор может быть интерпретирован, как фактор негативности долгого ожидания доходов от депозитных вкладов. Этот фактор угнетает остатки на всех счетах и их изменение на счетах депозитов любой срочности, и стимулирует рост процентных ставок краткосрочных депозитов.

Третий фактор имеет только две определяющие переменные: ставка процента по долгосрочным депозитам и ее изменение. Остальные переменные при этом факторе имеют незначительные нагрузки. Фактор определяется только депозитами счета 2635, поэтому он может быть интерпретирован, как привлекательность долгосрочных депозитов. Наибольшее негативное влияние этот фактор оказывает на изменения остатков депозитов на счете 2630. Влияние на остальные переменные несущественно (факторные нагрузки близки к нулевым значениям).

Рассмотрение значений факторных нагрузок уже дает ряд неожиданных результатов. Первый фактор – быстроты получения дохода от депозита, - не оказывает никакого влияния на уровень процентных ставок по краткосрочным депозитам (значение $-0,083$ близко к нулевому). Т.е. на формирование депозитных остатков по счету 2630 не влияет уровень процентной ставки. Лишь ее изменение вызывает прирост этих остатков. Действие первого фактора ведет не только к удержанию уровня процентных ставок по долгосрочным депозитам, но и снижает этот уровень. И логичное проявление существования

первого фактора состоит в том, что изменение остатков долгосрочных депозитов, благодаря этому, имеет знак минус ($a_{4,1} = -0,366$).

Второй фактор – негативности ожидания, – угнетает все показатели, кроме, как уже отмечалось, уровня процентных ставок по краткосрочным депозитам, и показателя изменения процентных ставок долгосрочных депозитов. Таким образом, фактор долгого ожидания стимулирует уровень ставок на счете 2630, делая более привлекательными депозиты с небольшими сроками, – и это с одной стороны. С другой стороны, долгое ожидание доходов может быть компенсировано приростом процентной ставки по долгосрочным депозитам, т.е. оговорка в депозитном договоре о возможности увеличения уровня процентной ставки по долгосрочным депозитам, может создать условия для формирования долгосрочных ресурсов.

Фактор привлекательности долгосрочных депозитов оказывает заметное и негативное влияние на изменение остатков на депозитных счетах. Однако этот фактор не стимулирует ни уровень, ни положительное изменение остатков долгосрочных депозитов. Другими словами, высокий уровень процентных ставок на счете 2635 и их рост не оказывают существенного влияния на величину долгосрочных депозитов и их положительное изменение. Этот вывод просто невозможно было бы сделать без привлечения факторного анализа.

Проанализируем поведение значений трех факторов в отдельные моменты времени. Значения факторов принимали величины, указанные в таблице 2.4.6.

Таблица 2.4.6 – Значения факторов депозитов в национальной валюте

Месяц	F_1	F_2	F_3
1	0,383	0,104	-0,241
2	0,410	0,111	-0,286
3	0,426	0,121	-0,112
4	0,402	0,123	0,301

Продолжение таблицы 2.4.6

5	-0,226	-0,222	0,244
6	0,054	-0,134	-0,153
7	0,064	-0,177	-0,381
8	0,076	-0,196	0,039
9	0,055	-0,261	-0,247
10	-0,237	-0,476	-0,288
11	0,235	-0,271	0,384
12	0,207	-0,317	0,474
13	0,107	-0,440	-0,106
14	0,332	-0,380	-0,032

Поведение значений факторов для разных моментов времени несет интересную информацию. Фактор быстроты получения доходов, практически для всех периодов времени оставался неизменно положительным, т.е. настроение клиентов поскорее получить доход является характерной чертой в поведении вкладчиков. Только в двух случаях желание скорого дохода отступало: в пятом и десятом месяце. Объяснение этому в пятом месяце простое: накануне произошло резкое повышение процентной ставки по долгосрочным депозитам, естественно, что на короткий промежуток времени стремление иметь доход почти сразу отступило на второй план.

В десятом месяце объяснение не такое очевидное: с основными показателями никаких резких изменений не было. Однако, процентная ставка в этом месяце перешагнула психологический барьер: она перешла из одного десятка в другой (была 20,11; а стала – 19,98). Видимо это обстоятельство послужило снижению интереса к депозитным операциям со стороны клиентов. Впрочем, уже в следующем месяце все вернулось на «круги своя».

Ключевым моментом для второго фактора стал опять же пятый месяц: начиная с него фактор негативности ожидания дохода меняет свой характер на противоположный. До четвертого месяца включительно его величины были

положительными, т.е. фактор работал в соответствии со своим названием – негативно. Начиная с пятого месяца, после резкого повышения ставки процента по счету 2635, величины второго фактора становятся отрицательными. Таким образом, негативное воздействие фактора заменяется его позитивным влиянием: продолжительное ожидание дохода по долгосрочным депозитам уже не смущает клиентов, они готовы ждать с примерно одинаковым усердием (примерно равное значение фактора для всех месяцев, начиная с пятого и заканчивая десятым).

Отрицательные значения третьего фактора в первые три месяца, говорят о том, что долгосрочные депозиты в этот период не были привлекательными. Только тогда, когда в четвертом месяце произошло резкое повышение ставки процента по долгосрочным депозитам, значение фактора приняло положительное и достаточно большое значение. В следующем месяце эта тенденция сохранилась, а дальше долгосрочные депозиты становились то привлекательными, то интерес к ним менялся на противоположный. Этому могут служить разные объяснения. К примеру, в восьмом месяце фактор привлекательности вернул себе положительное значение, чему, очевидно, послужило самое резкое снижение ставки процента по краткосрочным депозитам. Однако еще более резкое снижение этой ставки в следующем месяце не сделало долгосрочные депозиты более привлекательными: фактор вернул себе отрицательное значение. В одиннадцатом месяце фактор возвращает положительную величину вследствие, видимо, сократившегося падения процентной ставки по счету 2635.

Следовательно, несмотря на положительную динамику формирования остатков на долгосрочных депозитах, этот вид вкладов не пользуется постоянным вниманием. Интерес к ним носит неустойчивый характер, и об этом говорит поведение фактора привлекательности долгосрочных депозитов.

Если характеризовать воздействие всех трех факторов на процесс формирования остатков на депозитных счетах, то можно прийти к следующим выводам.

Фактор нетерпения поскорее получить доход имеет доминирующий характер, можно лишь на некоторое время отвлечь от него, но он неизменно возвращается к обычным величинам. Следствием действия этого фактора является возврат средств и уменьшение ресурсной базы. Это с одной стороны, и это – негативная часть существования данного фактора. Но с другой стороны, действие этого фактора поддерживает интерес к депозитным операциям, и благодаря этому поток приходящих клиентов будет постоянным.

Негативность длительного срока пребывания собственных денежных средств в других руках, в том числе и на банковском депозите, – как фактор, – может быть сломлена одноразовым резким повышением процентной ставки по долгосрочным депозитам. Этот фактор также может быть нейтрализован возможностью, даже теоретической, пересмотра ставки в сторону повышения по уже заключенным депозитным соглашениям. И в тоже время, третий фактор говорит о том, что к долгосрочным депозитам нет постоянного интереса, хотя это и наиболее важный источник формирования ресурсной базы КБ.

3. Факторный анализ многомерных временных рядов

3.1. Классический анализ временных рядов

Большинство статистических методов использует модели, в которых наблюдения в различные моменты времени предполагаются независимыми. Во многих приложениях зависимость между наблюдениями рассматривается как помеха. В планируемых экспериментах вводится рандомизация эксперимента, что позволяет обосновать анализ, проводимый без учета зависимости наблюдений. Однако разнообразные данные в коммерции, экономике, технике и естественных науках поступают в форме временных рядов, в которых наблюдения зависимы и характер этой зависимости интересен сам по себе. Временным рядом называется упорядоченная во времени последовательность

наблюдений. Совокупность существующих методов анализа таких рядов зависимых наблюдений называется анализом временных рядов.

Основной чертой, выделяющей этот анализ рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения. Если в некоторых задачах статистического анализа различных явлений значения переменных для разных моментов времени являются статистически независимыми, или это априорно предполагается, то во временных рядах они зависимы, и характер зависимости может определяться положением наблюдений в последовательности. Природа ряда и структура порождающего ряд процесса определяют порядок образования последовательности.

В экономической сфере имеется обширный круг явлений, которые интересно и важно изучать в их развитии и изменении во времени. С течением времени изменяются деловая активность, цены на различные материальные и финансовые активы, значения биржевых индексов, режим протекания того или иного производственного процесса, и тому подобное. Совокупность изменений какой-либо одной характеристики подобного рода в течение некоторого периода времени и представляет временной ряд. Это могут быть ежедневные величины индекса потребительских цен, или поквартальные данные о валовом национальном продукте.

Цели изучения временных рядов могут быть различными, и различные причины, по которым желательно исследовать временные ряды, могут определять выбор методов анализа явления. Различаются пять типов исследования временных рядов [28]:

- 1) на самом поверхностном уровне берут какой-либо ряд и строят простую математическую систему, которая *описывает* его поведение в сжатом виде;

- 2) проникая глубже, делают попытку *объяснить* поведение ряда с помощью других переменных и определить соотношение между наблюдениями

и некоторыми структурными законами поведения, т.е. для объяснения наблюдений в качестве гипотезы строится модель;

3) результаты анализа, проведенного на первых двух этапах, можно использовать для *прогнозирования* поведения ряда. В случае 1), даже когда ничего не известно об основном механизме, генерирующем ряд, исходят из предположения, что в системе существует достаточная инерция, чтобы гарантировать в будущем такое же поведение, как и в прошлом. Случай 2) характеризуется большим проникновением в область рассматриваемых причинно-следственных связей и можно делать проекции в будущее более уверенно;

4) в случае 2) может потребоваться *контроль* системы с целью выработки сигналов, предупреждающих о грядущих неблагоприятных событиях, причем сигнал может вырабатываться непосредственно либо самой моделью, либо на основании анализа того, что может случиться, если изменятся некоторые параметры модели, и эти изменения произойдут в реальности;

5) для еще более глубокого анализа экономического явления может возникнуть необходимость рассмотреть *совместное развитие* во времени нескольких переменных или, другими словами, исследуемая переменная может быть вектором наблюдений. И такой тип исследования представляет наибольший интерес.

Характеризуя временной ряд в целом, следует отметить, что в отличие от случайной выборки ряд следует рассматривать как смесь четырех компонент:

- 1) тренда или долгосрочного движения;
- 2) более или менее регулярных колебаний относительно тренда;
- 3) сезонной составляющей;
- 4) остатка или несистематического случайного эффекта.

Несомненно, ряд удобно представлять в виде суммы этих четырех компонент и одной из целей анализа является разложение ряда на эти составляющие для отдельного изучения. Для рядов экономических данных наличие тренда, вероятнее всего, может объясняться наличием постоянных

тенденций, однообразно действующих в одном направлении. Колебания относительно долгосрочного движения, скорее всего, происходят по совокупности объективных причин. И всем регулярным компонентам, в том числе и сезонной составляющей, присуще некоторое возмущение, порождаемое случайными событиями, появление которых предсказать не представляется возможным. Но то, что это имеет место в действительности и что эффекты от различных сил аддитивны, является предположением и имеет характер гипотезы, которую всегда следуют отклонить, если сконструированная модель плохо соответствует исходным данным.

Пусть наблюдаемым временным рядом являются значения некоторой переменной величины $z_1, z_2, \dots, z_T$, то есть имеется T чисел таких значений, наблюдаемых в равностоящие моменты времени. Эти моменты пронумерованы целыми числами $1, 2, \dots, T$. Отрицательные значения этих чисел означают соответствующие моменты в прошлом. Достаточно общей математической (статистической или вероятностной) моделью служит уравнение вида

$$z_t = f(t) + u_t, \quad t = 1, 2, \dots, T \quad (3.1.1)$$

В этой модели наблюдаемый ряд рассматривается как сумма некоторой полностью детерминированной последовательности $f(t)$, которую можно назвать систематической составляющей, и случайной последовательности u_t , подчиняющейся некоторому вероятностному закону. Эти компоненты наблюдаемого ряда ненаблюдаемы, они являются теоретическими величинами. Например, если фиксируется ежедневный объем продаж некоторой компании, то $f(t)$ может представлять планируемую величину реализации продукции, а u_t – особенности спроса данного дня, в эту же составляющую включаются искажения или ошибки в собираемой информации.

Рассмотрение временного параметра в модели согласуется с интуитивным представлением о том, что он собой представляет. Одно из таких представлений состоит в том, что время течет в одном направлении и не может быть обращено вспять. Другое – что события, близкие по времени, должны

быть сравнительно сильно связаны, а события, разделенные большим промежутком времени, не должны иметь сильной связи. Можно рассматривать различные варианты математической модели (3.1), в которых влияние времени может сказываться либо только на функции или последовательности $f(t)$, либо только на вероятностном процессе, определяющем случайную составляющую u_t , либо, наконец, на обеих этих компонентах.

Первоначально анализ временных рядов базировался на моделях, в которых влияние временного параметра проявлялось только в систематической составляющей. Эта ситуация называется классической [29], поскольку она в известной мере восходит к тем временам, когда Гаусс и его последователи развивали теорию и метод наименьших квадратов с целью применения их в астрономии и физике. В таких моделях предполагается, что течение времени никак не отражается на случайной составляющей. То есть предполагается, что математическое ожидание случайной составляющей тождественно равно нулю, дисперсия равна некоторой постоянной и что значения u_t в различные моменты времени некоррелированы. Такое определение приводит к тому, что всякую зависимость от времени приходится включать в систематическую составляющую $f(t)$.

Последовательность $f(t)$ может зависеть от некоторых неизвестных коэффициентов и от известных величин, меняющихся со временем. В этом случае ее называют функцией регрессии. Методы статистических выводов для коэффициентов функции регрессии оказываются полезными во многих областях статистики. Своеобразие методов, относящихся именно к временным рядам, состоит в том, что здесь исследуются те модели, в которых упомянутые выше величины, меняющиеся со временем, являются известными функциями t .

Анализ временных рядов имеет важное практическое значение во многих областях деятельности. А одной из важных практических целей является прогнозирование временных рядов. Использование доступных к моменту времени t наблюдений временного ряда для прогнозирования его значения в

некоторый момент времени в будущем $t + l$ может явиться основой для планирования в экономике и торговле, планирования выпуска продукции, складского контроля и контроля выпуска, управления и оптимизации промышленных процессов, решения социальных проблем и многих-многих других задач. Для всех этих задач существует необходимость в прогнозе вперед на интервал, называемым временем упреждения и зависящий от конкретной проблемы. К примеру, время упреждения в проблеме складского контроля определено как период, начинающийся с момента передачи заказа на пополнение склада поставщиком и длящийся до тех пор, пока заказ не доставлен на склад.

Предполагается, что наблюдения доступны в дискретные, равноотстоящие моменты времени. Например, в проблеме прогнозирования сбыта продукции сбыт z_t в текущем месяце t и сбыт в предыдущие месяцы могут быть использованы для прогноза сбыта с упреждением $l = 1, 2, \dots, 12$ месяцев.

Обозначим через $\hat{z}_t(l)$ сделанный в момент t прогноз сбыта z_{t+l} в некоторый момент $t + l$ в будущем, т.е. с упреждением l .

Функция $\hat{z}_t(l), l = 1, 2, \dots$, дающая в момент времени t прогнозы для всех будущих времен упреждения, называется прогнозирующей функцией в момент времени t [30]. Цель исследования временного ряда состоит в том, чтобы получить такую прогнозирующую функцию, у которой среднее значение квадрата отклонения $z_{t+l} - \hat{z}_t(l)$ – истинного значения от прогнозируемого значения является наименьшим для каждого упреждения l . Такой прогноз традиционно считается наилучшим.

В дополнение к вычислению наилучшего прогноза необходимо также указать его точность, чтобы, например, можно было бы оценить риск, связанный с решениями, основанными на прогнозировании. Точность прогноза может быть выражена вероятностными пределами по обе стороны от каждого прогнозного значения, т.е. являться следствием построения интервального

прогноза. Смысл такого прогнозного интервала состоит в том, что значение временного ряда, которое появится в соответствующее время, с заданной наперед вероятностью будет лежать внутри этого интервала.

Идея использования математических моделей для описания поведения физических объектов является общепризнанной. В частности, иногда удается получить модель, основанную на физических законах, что дает возможность вычислить точное значение какой-либо зависящей от времени величины в любой момент времени. Такая модель является детерминированной. Однако для большинства практических задач получить совершенно точное значение исследуемого признака нельзя из-за случайных возмущений, действующих на объект исследования. Например, остатки на депозитных счетах коммерческого банка на определенный момент, на которые оказывают влияние большое число различных факторов. Для таких объектов нельзя предложить детерминированную модель, допускающую точное вычисление будущего поведения объекта. Такая модель называется вероятностной или стохастической.

Модели временных рядов, необходимые для получения оптимального прогнозирования в действительности являются стохастическими. При этом необходимо различать стохастический процесс, с которым идентифицируется вероятностная модель, и наблюдаемый временной ряд [30, с.22]. Временной ряд $z_1, z_2, \dots, z_T$ из T последовательных наблюдений рассматривается как выборочная реализация из бесконечной популяции таких временных рядов, генерируемых стохастическим процессом.

Важным и одним из самых ранних классов стохастических моделей для описания временных рядов является класс стационарных моделей. Они основаны на предположении, что основные вероятностные характеристики стохастического процесса являются неизменными с течением времени, т.е. процесс остается в равновесии относительно постоянного среднего уровня. Однако в промышленности, коммерции, экономике и финансовой сфере, где прогнозирование имеет особо важное значение, многие временные ряды лучше

описываются нестационарными моделями, к которым, в частности, можно отнести модель экспоненциально взвешенных скользящих средних [31].

Укажем на основные виды классических моделей, используемых в анализе временных рядов [30]. Для этого необходимо ввести некоторые простые операторы.

Оператор сдвига назад B , определяемый как $Bz_t = z_{t-1}$, и тогда, соответственно, возведение оператора в степень m будет означать сдвиг назад на m шагов: $B^m z_t = z_{t-m}$.

Оператор сдвига вперед (обратная по отношению к оператору B) $F = B^{-1}$. Этот оператор задается как $Fz_t = z_{t+1}$, и, соответственно, $F^m z_t = z_{t+m}$.

Разностный оператор со сдвигом назад ∇ , который можно выразить через оператор B следующим образом: $\nabla z_t = z_t - z_{t-1} = (1 - B)z_t$.

Оператор суммирования S – обратный оператору ∇ , выражаемый как

$$\begin{aligned} Sz_t &= \sum_{j=0}^{\infty} z_{t-j} = z_t + z_{t-1} + z_{t-2} + \dots = (1 + B + B^2 + \dots)z_t = (1 - B)^{-1}z_t = \\ &= \nabla^{-1}z_t. \end{aligned}$$

Модель линейного фильтра. Эта модель исходит из предположения о том, что временные ряды, в которых последовательные значения сильно зависимы, следует рассматривать как генерируемые последовательностью независимых импульсов a_t . Эти импульсы – реализации случайных величин с фиксированным распределением, которое обычно предполагается нормальным с нулевым средним и дисперсией σ_a^2 . Такая последовательность случайных величин $a_t, a_{t+1}, a_{t+2}, \dots$ называется белым шумом.

Предполагается, что белый шум a_t можно трансформировать в процесс z_t при помощи линейного фильтра $\psi(B)$. Операция линейной фильтрации заключается в вычислении взвешенной суммы предыдущих наблюдений так, что стохастический процесс приобретает вид:

$$z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t. \quad (3.1.2)$$

В общем μ – параметр, определяющий «уровень» процесса, и

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

– линейный оператор, преобразующий a_t в z_t и называемый передаточной функцией фильтра.

Последовательность $\psi_1, \psi_2, \dots$, образованная весами, может быть конечной или бесконечной. Если эта последовательность является сходящейся, фильтр называется устойчивым, а процесс z_t будет стационарным. Параметр μ в этом случае – среднее значение, вокруг которого процесс варьирует. В других случаях z_t нестационарен, и μ не имеет какого-либо иного смысла, кроме как некоторой точки отсчета уровня процесса.

Модели авторегрессии. Так называемая модель авторегрессии является исключительно полезной стохастической моделью для описания некоторых часто встречающихся на практике рядов. В этой модели текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и импульса a_t . Обозначим значения стохастического процесса в равноотстоящие моменты времени $t, t-1, t-2, \dots$ как $z_t, z_{t-1}, z_{t-2}, \dots$. Пусть $\tilde{z}_t, \tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots$ будут отклонениями от μ , например $\tilde{z}_t = z_t - \mu$. Тогда

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t \quad (3.1.3)$$

называется процессом авторегрессии (АР) порядка p . В (3.1.3) $\phi_1, \phi_2, \dots, \phi_p$ – некоторые коэффициенты.

Название «авторегрессия» объясняется тем, что линейная модель вида

$$z = \phi_1 x_1 + \phi_2 x_2 + \dots + \phi_p x_p + a,$$

связывающая зависимую переменную z со множеством, так называемых независимых переменных $x_1, x_2, \dots, x_p$ плюс свободный член a , отождествляемый с ошибкой или случайным отклонением, называется уравнением регрессии; при этом говорят, что z регрессирует на $x_1, x_2, \dots, x_p$. В уравнении (3.1.3) переменная $\tilde{z}_t$ регрессирует на свои предшествующие значения, поэтому модель авторегрессирующая.

Если определить, что оператор авторегрессии порядка p имеет вид

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

то модель авторегрессии можно компактно записать в следующем виде:

$$\phi(B)\tilde{z}_t = a_t.$$

Эта модель содержит $p + 2$ неизвестных параметра: $\mu, \phi_1, \phi_2, \dots, \phi_p, \sigma_a^2$, которые на практике следует оценить по наблюдениям. Дополнительный параметр σ_a^2 — дисперсия белого шума.

Нетрудно заметить, что модель авторегрессии является частным видом модели линейного фильтра (3.1.2). Величину $\tilde{z}_{t-1}$ можно исключить из правой части (3.1.3) подстановкой из уравнения (3.1.3), считая ее зависимой переменной:

$$\tilde{z}_{t-1} = \phi_1 \tilde{z}_{t-2} + \phi_2 \tilde{z}_{t-3} + \dots + \phi_p \tilde{z}_{t-p-1} + a_{t-1}.$$

Аналогичным образом можно исключить $\tilde{z}_{t-2}$ и так далее, получив в результате бесконечный ряд из a . С помощью введенных операторов модель авторегрессии можно записать в следующем виде:

$$\tilde{z}_t = \psi(B)a_t,$$

где

$$\psi(B) = \phi^{-1}(B).$$

Процессы авторегрессии могут быть как стационарными, так и нестационарными. Чтобы процесс был стационарным, нужно выбрать ϕ так, чтобы веса $\psi_1, \psi_2, \dots$, в выражении $\psi(B) = \phi^{-1}(B)$ образовывали сходящийся ряд.

Модели скользящего среднего. Модель авторегрессии (3.1.3) выражает отклонение стохастического процесса $\tilde{z}_t$ в виде конечной взвешенной суммы p предыдущих отклонений процесса $\tilde{z}_{t-1}, \tilde{z}_{t-2}, \dots, \tilde{z}_{t-p}$ плюс случайный импульс a_t . А также эта модель выражает это же отклонение как бесконечную взвешенную сумму случайных импульсов.

Другим очень важным в описании наблюдаемых временных рядов типом моделей является так называемый процесс скользящего среднего. Пусть $\tilde{z}_t$

линейно зависит от конечного числа q предыдущих случайных импульсов a . Такой стохастический процесс

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3.1.4)$$

называется процессом скользящего среднего (СС) порядка q , где $\theta_1, \theta_2, \dots, \theta_q$ коэффициенты модели или веса, которые не обязательно должны быть положительными.

Если определить оператора скользящего среднего порядка q как

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

то модель скользящего среднего можно компактно представить в следующем виде:

$$\tilde{z}_t = \theta(B)a_t. \quad (3.1.5)$$

Она содержит $q + 2$ неизвестных параметра $\mu, \theta_1, \theta_2, \dots, \theta_p, \sigma_a^2$, которые на практике должны оцениваться по наблюдениям.

Смешанные модели авторегрессии – скользящего среднего. Для достижения большей гибкости в подгонке моделей к наблюдаемым временным рядам иногда целесообразно объединить в одной модели и авторегрессию, и скользящее среднее. Это приводит к комбинированной модели авторегрессии – скользящего среднего

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \dots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}, \quad (3.1.6)$$

или

$$\phi(B)\tilde{z}_t = \theta(B)a_t,$$

в которой имеется $p + q + 2$ неизвестных параметра:

$\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_p, \sigma_a^2$, оцениваемых по наблюдаемым данным.

На практике часто оказывается, что адекватное описание наблюдаемых временных рядов достигается при помощи моделей авторегрессии, скользящего среднего или комбинированной модели, в которых p и q не больше, а часто и меньше 2 [30].

Нестационарные модели. Многие ряды, практически встречающиеся в промышленности или торговле (например, биржевые цены), обнаруживают нестационарный характер и, в частности, не колеблются относительно фиксированного среднего. Тем не менее, их свойства могут быть в некотором смысле однородными. Например, хотя уровень относительно которого происходят флуктуации, может быть разным в разные времена, поведение рядов с учетом различий в уровне оказывается во многом сходным. Такой ряд может быть представлен обобщенным оператором авторегрессии $\varphi(B)$, в котором один или несколько нулей полинома $\varphi(B)$ (т.е. один или несколько корней уравнения $\varphi(B) = 0$) равны единице. В этом случае оператор $\varphi(B)$ имеет вид

$$\varphi(B) = \phi(B)(1 - B)^d,$$

где $\phi(B)$ – оператор стационарной авторегрессии, d – некоторое целое число. При этом обобщенная модель, описывающая однородный нестационарный процесс, имеет вид

$$\varphi(B)z_t = \phi(B)(1 - B)^d z_t = \theta(B)a_t,$$

т.е.

$$\phi(B)w_t = \theta(B)a_t, \quad (3.1.7)$$

где

$$w_t = \nabla^d z_t. \quad (3.1.8)$$

Таким образом, однородный нестационарный процесс может быть описан моделью, которая требует чтобы ∇^d -я разность процесса была стационарной. На практике d обычно равно 0, 1 или максимум 2.

Процесс, определенный (3.1.7) или (3.1.8), представляет собой эффективную модель для описания стационарных и нестационарных временных рядов и называется процессом авторегрессии – проинтегрированного скользящего среднего (АРПСС) порядка (p, d, q) . Процесс определен как

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3.1.9)$$

с $w_t = \nabla^d z_t$. Следует отметить, что если заменить w_t на $z_t - \mu$, то при $d = 0$ модель (3.9) будет содержать как частный случай и стационарную комбинированную модель (3.6), а также модели авторегрессии и скользящего среднего.

Если из выражения (3.8) выразить z_t , то получится следующее соотношение:

$$z_t = S^d w_t, \quad (3.1.10)$$

где S – оператор суммирования, определенный как

$$S w_t = \sum_{j=0}^{\infty} w_{t-j} = w_t + w_{t-1} + w_{t-2} + \dots$$

Таким образом, обобщенный процесс авторегрессии – проинтегрированного скользящего среднего можно получить из белого шума a_t с помощью трех операций фильтрации. Первый фильтр имеет вход a_t передаточную функцию $\theta(B)$ и выход e_t , где

$$e_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} = \theta(B) a_t. \quad (3.1.11)$$

Второй фильтр имеет вход e_t , передаточную функцию $\phi^{-1}(B)$ и выход w_t

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + e_t = \phi^{-1}(B) e_t. \quad (3.1.12)$$

И наконец, третий фильтр имеет вход w_t и выход z_t с передаточной функцией S^d .

Для описания сезонных временных рядов используется специальная форма модели (3.1.9).

Рассмотрим некоторые проблемы прогнозирования с использованием универсальной модели процесса авторегрессии – проинтегрированного скользящего среднего (АРПСС)

$$\varphi(B) z_t = \theta(B) a_t, \quad (3.1.13)$$

где оператор $\varphi(B)$ имеет вид следующего выражения: $\varphi(B) = \phi(B) \nabla^d$.

Рассмотрим прогноз значения $z_{t+l}, l \geq 1$ по отношению к текущему моменту времени t . О таком прогнозе можно сказать, что он делается в момент времени t с упреждением l .

Для этого проанализируем три явные формы представления модели. Наблюдение z_{t+l} , генерируемое процессом (3.1.13), можно выразить следующим образом:

1) непосредственно при помощи разностного уравнения

$$z_{t+l} = \varphi_1 z_{t+l+1} + \dots + \varphi_{p+d} z_{t+l-p-d} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q} + a_{t+l}; \quad (3.1.14)$$

2) как бесконечную взвешенную сумму текущего и предшествующих импульсов a_j

$$z_{t+l} = \sum_{j=-\infty}^{t+l} \psi_{t+l-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+l-j}, \quad (3.1.15)$$

где ψ_0 и веса ψ можно найти приравниванием коэффициентов при одинаковых степенях B правой и левой частей уравнения

$$\varphi(B)(1 + \psi_1 B + \psi_2 B^2 + \dots) = \theta(B). \quad (3.1.16)$$

3) как бесконечную взвешенную сумму предыдущих наблюдений плюс случайный импульс

$$z_{t+l} = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} + a_{t+l}. \quad (3.1.17)$$

Если $d \geq 1$, то

$$\tilde{z}_{t+l-1} = \sum_{j=1}^{\infty} \pi_j z_{t+l-j} \quad (3.1.18)$$

будет взвешенным средним, потому что сумма весов равна единице: $\sum_{j=1}^{\infty} \pi_j = 1$. Веса π можно получить из равенства

$$\varphi(B) = (1 - \pi_1 B - \pi_2 B^2 - \dots) \theta(B). \quad (3.1.19)$$

Обозначим прогноз величины z_{t+l} как $\hat{z}_t(l)$, при этом он является линейной функцией текущего и предшествующих наблюдений $z_t, z_{t-1}, z_{t-2}, \dots$. Поэтому он также будет линейной функцией текущего и предшествующих импульсов $a_t, a_{t-1}, a_{t-2}, \dots$.

Предположим, что наилучший прогноз существует и имеет вид

$$\hat{z}_t(l) = \psi_l^* a_t + \psi_{l+1}^* a_{t-1} + \psi_{l+2}^* a_{t-2} + \dots,$$

где веса $\psi_l^*, \psi_{l+1}^*, \dots$ должны быть определены. Тогда с учетом (3.1.14) среднеквадратическая ошибка прогноза равна

$$\mathcal{M}\{z_{t+l} - \hat{z}_t(l)\}^2 = (1 + \psi_1^2 + \dots + \psi_l^2)\sigma_a^2 + \sum_{j=0}^{\infty}(\psi_{l+j} - \psi_{l+j}^*)^2 \sigma_a^2. \quad (3.1.20)$$

В последнем выражении $\mathcal{M}\{\cdot\}$ – символ математического ожидания.

Ошибка прогноза может быть минимизирована приравниванием $\psi_{l+j} = \psi_{l+j}^*$. Этот вывод является частным случаем более общих результатов теории прогнозирования, развитой Колмогоровым [32]. В результате имеет место соотношение [30]:

$$z_{t+l} = (a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1}) + (\psi_l a_t + \psi_{l+1} a_{t-1} + \dots), \quad (3.1.21)$$

$$z_{t+l} = e_t(l) + \hat{z}_t(l), \quad (3.1.22)$$

где $e_t(l)$ – ошибка прогноза $\hat{z}_t(l)$ с упреждением l .

Отсюда вытекает ряд важных выводов. Обозначим условное математическое ожидание z_{t+l} при условии, что все z до момента t известны через $\mathcal{M}_t\{z_{t+l}\}$. Тогда

$$\hat{z}_t(l) = \psi_l a_t + \psi_{l+1} a_{t-1} + \dots = \mathcal{M}_t\{z_{t+l}\}. \quad (3.1.23)$$

Это значит, что прогноз с минимальной среднеквадратической ошибкой в момент t с упреждением l есть условное математическое ожидание z_{t+l} в момент t . Когда $\hat{z}_t(l)$ рассматривается как функция l при фиксированном t она называется прогнозирующей функцией для момента t .

Ошибка прогноза для упреждения l равна

$$e_t(l) = a_{t+l} + \psi_1 a_{t+l-1} + \dots + \psi_{l-1} a_{t+1}. \quad (3.1.24)$$

Так как математическое ожидание ошибки равно нулю: $\mathcal{M}_t\{e_t(l)\}$, прогноз будет несмещенный.

Дисперсия ошибки прогноза $V(l)$ равна

$$V(l) = \text{var}\{e_t(l)\} = (1 + \psi_1^2 + \dots + \psi_{l-1}^2)\sigma_a^2. \quad (3.1.25)$$

Еще один важный вывод состоит в том, что не только $\hat{z}_t(l)$ есть прогноз величины z_{t+l} с минимальной среднеквадратической ошибкой, но и любая линейная комбинация прогнозов $\sum_{l=1}^L w_l \hat{z}_t(l)$, где L – число таких прогнозов, w – некоторые веса этих прогнозов, – также является прогнозом с минимальной

среднеквадратической ошибкой соответствующей линейной комбинации будущих наблюдений: $\sum_{l=1}^L w_l z_{t+l}$.

Например, пусть с помощью выражения (3.1.23) по ежемесячным данным получен прогноз сбыта товара на один, два и три месяца вперед: $\hat{z}_t(1), \hat{z}_t(2), \hat{z}_t(3)$. Тогда $\hat{z}_t(1) + \hat{z}_t(2) + \hat{z}_t(3)$ – это прогноз с минимальной среднеквадратической ошибкой сбыта товара $z_{t+1} + z_{t+2} + z_{t+3}$ в течение следующего квартала.

Остаточные ошибки как ошибки прогноза на один шаг вперед определяются на основании выражения (3.1.19), и имеют вид

$$e_t(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1}. \quad (3.1.26)$$

Следовательно, остаточные ошибки, генерирующие процесс, ошибки, которые вводятся как множество независимых случайных величин или импульсов, тождественны ошибкам прогноза на шаг вперед.

Отсюда следует, что для прогноза с минимальной среднеквадратической ошибкой, ошибки прогноза на шаг вперед должны быть некоррелированными. Это утверждение выглядит вполне логичным, поскольку в противном случае, т.е. если бы они были коррелированными, то ошибка прогноза на шаг вперед a_{t+1} могла быть предсказана на основании известных ошибок прогноза $a_t, a_{t-1}, a_{t-2}, \dots$. И в этом случае сумма прогноза ошибки и самого прогноза был бы лучшим прогнозом, что противоречит утверждению о единственности лучшего прогноза.

Однако следует отметить, что ошибки прогноза $e_t(l)$ и $e_t(l+j)$, сделанные для разных упреждений с одного и того же момента времени t , все-таки коррелированы, если эти моменты l и j времени значительно отстоят от текущего момента времени t . Одно из следствий этого состоит в том, что часто функции прогноза лежат либо целиком выше, либо целиком ниже фактических значений ряда.

То, что прогноз с минимальной среднеквадратической ошибкой прогноза $\hat{z}_t(l)$ для упреждения l – это условное математическое ожидание $\mathcal{M}_t\{z_{t+l}\}$

случайной величины z_{t+l} в момент времени t , позволяет записать выражения для прогноза любым из трех различных способов, соответствующих трем способам представления модели: (3.14), (3.15) и (3.17). Для $l > 0$ имеется три различных способа выражения прогнозов.

Прогнозы, полученные из разностного уравнения. Переходя в (3.14) к условным математическим ожиданиям в момент времени t , получаем

$$\begin{aligned}\mathcal{M}_t\{z_{t+l}\} &= \hat{z}_t(l) \\ &= \varphi_1 \mathcal{M}_t\{z_{t+l-1}\} + \dots + \varphi_{p+d} \mathcal{M}_t\{z_{t+l-p-d}\} - \theta_1 \mathcal{M}_t\{a_{t+l-1}\} - \dots - \\ &- \theta_q \mathcal{M}_t\{a_{t+l-q}\} + a_{t+l}.\end{aligned}\quad (3.1.27)$$

Прогноз в проинтегрированном виде. Пользуясь (3.15), можно получить выражение для прогноза

$$\begin{aligned}\mathcal{M}_t\{z_{t+l}\} &= \hat{z}_t(l) = \psi_1 \mathcal{M}_t\{a_{t+l-1}\} + \dots + \psi_{l-1} \mathcal{M}_t\{a_{t+1}\} + \psi_l \mathcal{M}_t\{a_t\} + \\ &+ \psi_{l+1} \mathcal{M}_t\{a_{t-1}\} + \dots + a_{t+1},\end{aligned}\quad (3.1.28)$$

Прогноз, как взвешенное среднее предшествующих наблюдений и прогнозов, сделанных в тот же момент с меньшими упреждениями. Наконец, переходя в (3.1.17) к условным математическим ожиданиям, можно получить третью форму выражения прогноза:

$$\mathcal{M}_t\{z_{t+l}\} = \hat{z}_t(l) = \sum_{j=1}^{\infty} \pi_j \mathcal{M}_t\{z_{t+l-j}\} + \mathcal{M}_t\{a_{t+l}\}.\quad (3.1.29)$$

Нужно отметить, что прогноз с минимальной среднеквадратической ошибкой определен через условное математическое ожидание

$$\mathcal{M}_t\{z_{t+l}\} = \mathcal{M}_t\{z_{t+l} | z_t, z_{t-1}, \dots\},$$

для которого теоретически требуется знание всех прошлых z до бесконечности. Однако условие обратимости, наложенное на общую модель АРПСС, требует, чтобы веса π образовывали сходящийся ряд. Отсюда следует, что для вычисления прогноза с заданной точностью, начиная с некоторого k , влиянием $z_{t-j}, j > k$, можно пренебречь. На практике веса π обычно затухают довольно быстро так, что какое бы представление модели ни использовалось в вычислениях, лишь небольшая часть ряда $z_t, z_{t-1}, \dots, z_{t-k}$ необходима для вычисления прогноза с достаточной точностью.

При вычислении условных математических ожиданий, входящих в выражения (3.27), (3.28), (3.29), следует учесть, что если j – неотрицательное числа, то

$$\begin{aligned}\mathcal{M}_t\{z_{t-j}\} &= z_{t-j}, j = 0, 1, 2, \dots, \\ \mathcal{M}_t\{z_{t+j}\} &= \hat{z}_t(j), j = 0, 1, 2, \dots, \\ \mathcal{M}_t\{a_{t-j}\} &= a_{t-j} = z_{t-j} - \tilde{z}_{t-j-1}(1), j = 0, 1, 2, \dots, \\ \mathcal{M}_t\{a_{t+j}\} &= 0, j = 0, 1, 2, \dots.\end{aligned}\quad (3.1.30)$$

Следовательно, чтобы получить прогноз $\hat{z}_t(l)$, необходимо выразить модель z_{t+l} с помощью любого из трех рассмотренных выше явных представлений и поступать с членами в правой части согласно следующим правилам:

- члены z_{t-j} , ($j = 0, 1, 2, \dots$), известные уже к моменту t , оставить без изменения;
- члены z_{t+j} , ($j = 0, 1, 2, \dots$), еще не известные, заменить их прогнозами $\hat{z}_t(j)$ на момент времени t ;
- члены a_{t-j} , ($j = 0, 1, 2, \dots$), уже известные, определить по выражению $z_{t-j} - \tilde{z}_{t-j-1}(1)$;
- члены уравнений вида a_{t+j} , ($j = 0, 1, 2, \dots$), еще не известные, заменить нулями.

Непосредственный выбор модели, по которой осуществляется прогноз, зависит от особенностей поставленной в исследовании задачи. На практике же чаще всего удобнее пользоваться непосредственно представлением модели разностным уравнением (3.1.14).

Пример [30]. Ряд значений биржевых цен на акции компании IBM к концу торгов хорошо описываются моделью

$$(1 - 0,8B)(1 - B)z_{t+l} = a_{t+l},$$

т.е.

$$(1 - 1,8B + 0,8B^2)z_{t+l} = a_{t+l},$$

или

$$z_{t+l} = 1,8z_{t+l-1} - 0,8z_{t+l-2} + a_{t+l}.$$

Прогнозы в момент времени t будут иметь вид

$$\hat{z}_t(1) = 1,8z_t - 0,8z_{t-1},$$

$$\hat{z}_t(2) = 1,8\hat{z}_t(1) - 0,8z_t,$$

$$\hat{z}_t(l) = 1,8\hat{z}_t(l-1) - 0,8\hat{z}_t(l-2), l = 3, 4, 5, \dots$$

Видно, что прогнозы легко вычисляются рекуррентным способом, начиная с $\hat{z}_t(1), \hat{z}_t(2), \dots$.

Классический анализ временных рядов рассматривает лишь одну случайную величину. Рассмотрение совокупности переменных, одновременно наблюдаемых в соответствующие моменты времени, выходит за рамки классических или традиционных представлений. А такой подход является достаточно актуальным, поскольку на значение переменной оказывает влияние не только время, как некоторый фактор, а также и значения других переменных, от которых данная переменная может зависеть вследствие объективных причинно-следственных связей.

В данном разделе рассмотрен так называемый параметрический подход к анализу временных рядов. Помимо него существует также спектральный анализ временных рядов или непараметрический подход. Представление наблюдаемых значений, как некоторого сигнала в форме ряда Фурье, также может быть использовано для целей прогнозирования, однако, чаще всего исследование гармоник, на которые раскладывается реальный ряд, служит для анализа и объяснения апостериорных значений переменной.

3.2. Основные методы в анализе многомерных временных рядов

В анализе временных рядов исторически имели место два довольно обособленных направления в исследовании временных рядов: частотный или гармонический подход и направление, связанное с анализом зависимости от времени.

Наиболее полный обзор методов анализа многомерных временных рядов с применением частотного подхода приведен у Д. Бриллинджера [33]. Как уже отмечалось, практическое использование методов многомерных временных рядов для анализа экономических проблем не нашло достаточного распространения по ряду причин. Важным условием применения этих методов является предположение об инвариантности по отношению к временным сдвигам: доля значений анализируемого признака, попавших в некоторый временной интервал, должна быть примерно такой же для другого временного интервала той же длины, но имеющего некоторый временной сдвиг по отношению к первоначальному интервалу.

Типичные физические эксперименты обладают свойством временной инвариантности. Например, для многих практических целей неважно, в какой именно день начата серия измерений силы тяжести; если рассмотреть ряд изменений температуры в течение десятков лет, можно убедиться, что он обладает свойством инвариантности. Однако поведение рядов, имеющих отношение к экономике и связанных с социально-экономическими процессами, не обладают свойством инвариантности по отношению ко времени. Объяснением этому служит достаточно очевидное предположение: люди извлекают уроки из прошлого и соответственно меняют свое поведение. Поэтому ряды, имеющие отношение к человеческой деятельности, вообще говоря, не обладают временной инвариантностью [33, с. 13].

Воспроизведем основную модель и основные результаты в многомерном анализе временных рядов, связанных с частотным подходом. Рассматривается $(r + s)$ -мерный временной стационарный ряд

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix}, \quad (3.2.1)$$

$t = 0, \pm 1, \pm 2, \dots$, составленный из r -мерного ряда $X(t)$ и s -мерного ряда $Y(t)$.

Обозначим математические ожидания компонент ряда

$$\mathcal{M}\{X(t)\} = \mu_X,$$

$$\mathcal{M}\{Y(t)\} = \mu_Y, \quad (3.2.2)$$

и ковариации

$$\begin{aligned} \mathcal{M}\{(X(t+u) - \mu_X)(X(t+u) - \mu_X)^T\} &= \Sigma_{XX}(u), \\ \mathcal{M}\{(X(t+u) - \mu_X)(Y(t+u) - \mu_Y)^T\} &= \Sigma_{XY}(u), \\ \mathcal{M}\{(Y(t+u) - \mu_Y)(Y(t+u) - \mu_Y)^T\} &= \Sigma_{YY}(u), \end{aligned} \quad (3.2.3)$$

$u = 0, \pm 1, \pm 2, \dots$, а также спектральные плотности второго порядка

$$\begin{aligned} f_{XX}(\lambda) &= \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \Sigma_{XX}(u) e^{-i\lambda u}, \\ f_{XY}(\lambda) &= \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \Sigma_{XY}(u) e^{-i\lambda u}, \\ f_{YY}(\lambda) &= \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \Sigma_{YY}(u) e^{-i\lambda u}, \end{aligned} \quad (3.2.4)$$

где λ – частота гармоники ($-\infty < \lambda < \infty$), $i = \sqrt{-1}$ – мнимая единица.

Задача анализа временных рядов состоит в выборе такого s -мерного вектора $\mathbf{c}$ и такого $s \times r$ -фильтра $\mathbf{A}\{u\}$, чтобы ряд

$$\mathbf{c} + \sum_{u=-\infty}^{\infty} \mathbf{A}(t+u)\mathbf{X}(u) \quad (3.2.5)$$

был в некотором смысле близок к $\mathbf{Y}(t)$.

Получим оценки $\mathbf{c}$ и $\mathbf{A}\{u\}$, основанные на конечных выборках значений $\mathbf{X}(t)$ и $\mathbf{Y}(t)$, $t = 0, 1, \dots, T-1$.

Пусть ряд (3.2.1) имеет выборочное среднее значение

$$\begin{pmatrix} \mathbf{c}_X \\ \mathbf{c}_Y \end{pmatrix}, \quad (3.2.6)$$

и выборочную ковариационную матрицу

$$\begin{pmatrix} \mathbf{C}_{XX} & \mathbf{C}_{XY} \\ \mathbf{C}_{YX} & \mathbf{C}_{YY} \end{pmatrix}. \quad (3.2.7)$$

Укажем на такой s -компонентный вектор $\mathbf{c}$ и такую $(s \times r)$ матрицу $\mathbf{A}$, которые минимизируют $(s \times s)$ матрицу

$$\mathcal{M}\{(\mathbf{Y} - \mathbf{c} - \mathbf{AX})(\mathbf{Y} - \mathbf{c} - \mathbf{AX})^T\}. \quad (3.2.8)$$

При этом под минимумом по отношению к матрицам понимается упорядочение $A \geq B$, означающее, что матрица $A - B$ неотрицательно определена.

Пусть задан $(r + s)$ -мерный вектор

$$\begin{pmatrix} X \\ Y \end{pmatrix}$$

со средним (3.2.6) и ковариационной матрицей (3.2.7) и матрица C_{XX} – невырождена, тогда матрицу (3.2.8) минимизируют следующие величины

$$c = c_Y - C_{YX}C_{XX}^{-1}c_X \quad (3.2.9)$$

и

$$A = C_{YX}C_{XX}^{-1}. \quad (3.2.10)$$

Соответствующее минимальное значение равно

$$C_{YY} - C_{YX}C_{XX}^{-1}C_{XY}. \quad (3.2.11)$$

При этом случайный вектор

$$c_Y + C_{YX}C_{XX}^{-1}(X - c_X) \quad (3.2.12)$$

называется наилучшим линейным прогнозом Y , основанным на X .

Величины (3.2.9) и (3.2.10) доставляют также минимум детерминанту, следу, диагональным элементам и собственным значениям матрицы (3.2.8).

Модель многомерного временного ряда (3.2.5) предполагает наличие детерминированного или стохастического ряда $X(t)$, который, по аналогии с классическим регрессионным анализом, интерпретируется как матрица значений независимых переменных. Тем самым эта модель является обобщением модели классического регрессионного анализа на случай зависимости экзогенных переменных от времени и на случай их стохастичности. Таким образом, этот подход аппроксимации исследуемого ряда с помощью фильтрации другого стационарного ряда и данная модель не предполагают наличия авторегрессии изучаемой совокупности переменных. И при этом также не рассматриваются проблемы корреляции между переменными результирующего ряда $Y(t)$.

Однако существует подход к решению проблем аппроксимации ряда посредством фильтрации не другого, а того же самого ряда. Рассмотрим основные особенности такого подхода.

Пусть имеется векторный ряд $X(t), t = 0, \pm 1, \pm 2, \dots$ с r компонентами. Его математическое ожидание для всех моментов времени равно

$$\mathcal{M}\{X(t)\} = \mu_X. \quad (3.2.13)$$

Этот ряд имеет абсолютно суммируемую автоковариационную функцию

$$\mathcal{M}\{(X(t+u) - \mu_X)(X(t+u) - \mu_X)^T\} = \Sigma_{XX}(u), \quad (3.2.14)$$

и матрицу спектральной плотности

$$f_{XX}(\lambda) = \frac{1}{2\pi} \sum_{u=-\infty}^{\infty} \Sigma_{XX}(u) e^{-i\lambda u}. \quad (3.2.15)$$

Чтобы пояснить основную идею этого подхода, воспользуемся аналогией из области передачи данных. Допустим, необходимо передать по каналам связи из одного пункта в другой значения величин $X(t)$, но при этом распоряжении имеется только $l \leq r$ каналов. Предположим, что в процессе передачи ряда данных $X(t)$ по q каналам связи можно описать как фильтрацию, в результате которой получается l -компонентный ряд

$$\Psi(t) = \sum_u B(t-u)X(u), \quad t = 0, \pm 1, \pm 2, \dots, \quad (3.2.16)$$

где $B(t-u)$ – $q \times r$ -матричный фильтр. Пусть на выходе каналов принимается ряд, который выступает в качестве оценки для $X(t)$:

$$X^*(t) = c + \sum_u D(t-u)\Psi(u), \quad (3.2.17)$$

где c – вектор с r компонентами,

$D(t-u)$ – $r \times q$ -фильтр.

Необходимо выбрать вектор c и фильтры $B(t-u)$, $D(t-u)$ так, чтобы ряд $X^*(t)$ был близок к ряду $X(t)$.

Эта задача может рассматриваться и как отыскание способа построить такой l -мерный ряд $\Psi(t)$, который несет значительную часть информации об исходном ряде $X(t)$. И в этой связи Бриллинджер [33, с. 363] делает такое замечание: «Показатели применяются для того, чтобы измерить изменения величины, не доступной прямому наблюдению, но о которой известно, что она оказывает определенное воздействие на другие величины, непосредственно наблюдаемые. Причем хотя само это воздействие вызывает либо одновременное увеличение, либо уменьшение всех наблюдаемых величин, его эффект маскируется действием многих других причин, по-разному влияющих на отдельные наблюдаемые».

Поэтому введенный ряд $\Psi(t)$ может выступать в роли ряда показателей, описывающих воздействие на $X(t)$ со стороны некоторого скрытого от наблюдателя ряда. При этом предполагается, что ряд $\Psi(t)$ наилучшим среди мерных рядов образом позволяет восстановить $X(t)$ с помощью линейных операций, инвариантных во времени.

Если ввести ряд $\varepsilon(t)$, описывающий ошибку или искажение передаваемой по каналам связи информации:

$$\varepsilon(t) = X(t) - X^*(t); \quad (3.2.18)$$

тогда получим

$$X(t) = c + \sum_u D(t-u)\Psi(u) + \varepsilon(t), t = 0, \pm 1, \pm 2, \dots \quad (3.2.19)$$

Ряд $X(t)$ оказывается выраженным как профильтрованный вариант ряда $\Psi(t)$, имеющего меньшую размерность, плюс ошибка.

Такие модели находят применение, к примеру, в сейсмологии: пусть $\Psi(u)$ представляет собой ряд импульсов, вызванных q землетрясениями, одновременно происходящих в разных местах. Ряд $X(t)$ представляет собой сигналы, принятые r сейсмографами, и $D(t)$ описывает переходные явления в земной коре, связанные с землетрясениями. Сейсмологов на основании фиксируемой информации о ряде $X(t)$ и априорных знаниях о поведении земной коры во время землетрясений $D(t)$, интересуют свойства ряда $\Psi(u)$.

Укажем на возможность получения оценок параметров модели (3.2.17), (3.2.16) на основании конечной выборки наблюдений. Пусть имеется X - случайная матрица с r компонентами и некоторым числом наблюдений, имеющая среднее значение c_X и ковариационную матрицу C_{XX} . Поставим задачу одновременной минимизации всех собственных значений симметричной матрицы

$$\mathcal{M}\{(X - c - DBX)(X - c - DBX)^T\} \quad (3.2.20)$$

за счет выбора вектора c с r компонентами, $(q \times r)$ -матрицы B и $(r \times q)$ -матрицы D . Определив соответствующие величины c, B, D , можно убедиться,

что они доставляют минимум также монотонным функциям от собственных чисел матрицы (3.2.20) таким, как след, детерминант и диагональные элементы.

Поскольку любая $r \times r$ матрица A ранга $q \leq r$ может быть представлена в виде произведения DB , в котором матрица B имеет размер $q \times r$, а матрица $C - r \times q$, то, тем самым, отыскивая B и C , будет одновременно найдена матрица A ранга, не превосходящего q , которая минимизирует собственные числа матрицы [33, с. 98]

$$\mathcal{M}\{(X - c - AX)(X - c - AX)^T\}. \quad (3.2.21)$$

Доказано [33, с. 364], что все собственные числа матрицы (3.2.20) минимальны, когда параметры модели имеют вид

$$B = \begin{pmatrix} q_1^T \\ \vdots \\ q_q^T \end{pmatrix}, \quad (3.2.22)$$

$$D = (q_1 \quad \dots \quad q_q) = B^T \quad (3.2.23)$$

и

$$c = c_X - DBc_X, \quad (3.2.24)$$

где q_j^T есть j -й собственный вектор матрицы C_{XX} , $j = 1, 2, \dots, r$. Если собственные значения матрицы C_{XX} обозначить как λ_j , $j = 1, 2, \dots, r$, то при выборе значений параметров модели, равных (3.2.22), (3.2.23) и (3.2.24), матрица (3.2.20) примет наименьшее значение, равное величине

$$\sum_{j>q} \lambda_j q_j q_j^T. \quad (3.2.25)$$

Величина

$$\psi_j = q_j^T X \quad (3.2.26)$$

называется j -й главной компонентой матрицы X , $j = 1, 2, \dots, r$. Таким образом, в модели (3.2.19) ряд $X(t)$ определяется посредством своих q главных компонент.

Поскольку собственные вектора симметричных матриц – ортогональны, то и главные компоненты, определяемых в анализе многомерных временных

рядов как (3.2.26), также являются ортогональными по отношению друг к другу.

Таким образом, при частотном подходе, делающем упор на исследование спектральных характеристик многомерных временных рядов, наблюдаемый ряд аппроксимируется с помощью своих главных компонент. Следует отметить, что в отличие от классического факторного анализа, в котором главной компонентой матрицы считается ее собственные векторы, в анализе многомерных временных рядов главная компонента – это произведение соответствующего собственного вектора на саму матрицу.

В подходе, связанном с анализом зависимости от времени, распространены те же типы моделей, что и в одномерном случае, а именно: авторегрессия, модель скользящего среднего и смешанная модель. Схемы построения моделей имеют сходную с одномерным вариантом структуру конструкции зависимостей. Так для авторегрессии в многомерном случае имеется r рядов и у каждого из них значение в момент времени t связано не только с его собственными значениями в предыдущие моменты времени, но и с предшествующими значениями других рядов. В каждом из r соотношений имеется остаточный член ε , который может быть разным для различных рядов. В таком случае для описания авторегрессионных рядов порядка k необходимо $p + kp^2$ констант-коэффициентов модели, даже если ε независимы и неавтокоррелированы, что, очевидно, приводит к чрезмерному количеству констант, когда p и k большие.

Для формирования общего вида модели рассмотрим, к примеру, три ряда $x_{1,t}, x_{2,t}, x_{3,t}$. Тогда полностью уравнения можно записать в виде:

$$\begin{cases} \alpha_{01}x_{1,t} + \alpha_{11}x_{1,t-1} + \alpha_{21}x_{2,t-1} + \alpha_{31}x_{3,t-1} = \varepsilon_{1t}; \\ \alpha_{02}x_{2,t} + \alpha_{12}x_{1,t-1} + \alpha_{22}x_{2,t-1} + \alpha_{32}x_{3,t-1} = \varepsilon_{2t}; \\ \alpha_{03}x_{3,t} + \alpha_{13}x_{1,t-1} + \alpha_{23}x_{2,t-1} + \alpha_{33}x_{3,t-1} = \varepsilon_{3t}; \end{cases} \quad (3.2.27)$$

или, используя оператор сдвига назад,

$$\begin{cases} (\alpha_{01} + \alpha_{11}B)x_{1,t} + \alpha_{21}Bx_{2,t} + \alpha_{31}Bx_{3,t} = \varepsilon_{1t}; \\ \alpha_{12}Bx_{1,t} + (\alpha_{02} + \alpha_{22}B)x_{2,t} + \alpha_{32}Bx_{3,t} = \varepsilon_{2t}; \\ \alpha_{13}Bx_{1,t} + \alpha_{23}Bx_{2,t} + (\alpha_{03} + \alpha_{33}B)x_{3,t} = \varepsilon_{3t}. \end{cases} \quad (3.2.28)$$

В матричной форме система уравнений (3.2.28) принимает вид:

$$(A_0 + A_1B)x_t = \varepsilon_t, \quad (3.2.29)$$

где матрицы имеют вид

$$A_0 = \begin{pmatrix} \alpha_{01} & 0 & 0 \\ 0 & \alpha_{02} & 0 \\ 0 & 0 & \alpha_{03} \end{pmatrix};$$

$$A_1 = \begin{pmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & \alpha_{22} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{pmatrix};$$

векторы – матрицы столбцы: x_t – вектор наблюдаемых рядов

$$x_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \\ x_{3,t} \end{pmatrix};$$

ε_t – вектор случайных отклонений или остатков:

$$\varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \varepsilon_{3,t} \end{pmatrix}.$$

Уравнение (3.2.29) представляет собой модель авторегрессии первого порядка трех рядов исследуемых переменных.

Модель авторегрессии второго порядка, по аналогии с (3.2.29) можно записать в следующем виде:

$$(A_0 + A_1B + A_2B^2)x_t = \varepsilon_t, \quad (3.2.30)$$

где матрица коэффициентов A_2 имеет вид, аналогичный матрице коэффициентов A_1 , а оператор B^2 производит сдвиг на два шага назад.

И общую авторегрессионную модель произвольного порядка можно записать как

$$A(B)x_t = \varepsilon_t, \quad (3.2.31)$$

где

$$A(B) = \sum_{m=0}^p A_m B^m, \quad (3.2.32)$$

p – порядок авторегрессии и A являются матрицами коэффициентов размерности $r \times r$.

Формальное решение разностного уравнения (3.2.31) можно записать в виде (систематические члены предполагаются затухающими)

$$x_t = A^{-1}(B)\varepsilon_t. \quad (3.2.33)$$

Обратная матрица A^{-1} в общем случае будет равна отношению двух полиномов от B . Разложение единицы, деленной на знаменатель, и умножение на числитель дает бесконечный ряд постоянных коэффициентов, имеющих характер весов при усреднении ε на бесконечном интервале.

Аналогично через матрицы коэффициентов, обозначаемых как M , можно записать процесс скользящего среднего, а именно:

$$x_t = M(B)\varepsilon_t, \quad (3.2.34)$$

где

$$M(B) = \sum_{m=0}^p M_m B^m. \quad (3.2.35)$$

Смешанный процесс многомерного временного ряда имеет вид:

$$A(B)x_t = M(B)\varepsilon_t. \quad (3.2.36)$$

Предположим, что все ε имеют нулевое математическое ожидание и дисперсию σ^2 . Тогда ковариационной матрицей между наблюдениями, расположенных на расстоянии в m шагов друг от друга, будет матрица

$$C_m = \mathcal{M}\{x_t x_{t-m}^T\} \quad (3.2.37)$$

и после подстановки выражения (3.2.34) для процесса скользящего среднего в выражение для ковариационной матрицы (3.2.37) оказывается, что $C_m = \sigma^2 \sum_{s} M_{m+s} M_s^T =$ коэфф. при B^m в $(\sum M_j B^j)(\sum M_j^T B^{-j})$, (3.2.38)

Аналогично для авторегрессионного процесса

$$C_m = \text{коэфф. при } B^m \text{ в } A^{-1}(B)(A^{-1}(B^{-1}))^T, \quad (3.2.39)$$

а для смешанного процесса

$$C_m = \text{коэфф. при } B^m \text{ в } A^{-1}(B)M(B)M^T(B^{-1})(A^{-1}(B^{-1}))^T. \quad (3.2.40)$$

Как заметил М. Кендэл [28, с. 145]: «Решить эти уравнения можно, но сделать это нелегко», к тому же эти уравнения могут иметь не единственное

решение, что порождает проблему неидентифицируемости. Корень этой проблемы, как отмечал М. Кендэл [28, с. 145], лежит в скользящем среднем, а не в авторегрессионной части модели. При этих обстоятельствах он рекомендует отказаться от компоненты, представленной в модели скользящим средним.

Помимо частотного и временного подходов к анализу многомерных временных рядов, широкое распространение получил, условно говоря, смешанный подход, известный под названием «Фильтр Калмана». Модели фильтрации (а модель авторегрессии, как отмечалось выше, может быть сведена к таковой), служащие для выделения полезной информации из шума, основаны на методах Фурье и в соответствии с этим предполагали стационарность или нечто, достаточно близкое к ней. Подобные методы приводят к описанию фильтра в терминах частотной характеристики, и для того, чтобы получить импульсную переходную функцию, необходим следующий шаг – разложение Фурье. При этом исследуемые признаки должны иметь характер непрерывных величин, что достаточно удобно для построения фильтра с применением аналоговых вычислительных машин. Однако при использовании современной вычислительной техники частотная характеристика не является удобным средством описания фильтра.

Еще два фактора подводят к необходимости модификации модели многомерных временных рядов. Во-первых, на практике спектры имеют не столь общий характер, как это может показаться из научной литературы [34]. Более того, они либо даются исходной физической теорией (кроме констант, подлежащих оценке), либо получаются в результате оценки по прошлым данным. Второй фактор связан с возможностью нестационарности. Одной из причин нестационарности может быть тот факт, что для приближения любой системы к стационарному состоянию требуется определенное время. Если же система наблюдается с момента, когда в ней происходят некоторые переходные процессы, то стационарная модель будет неподходящей. К тому же, системе по ее физической природе может быть присуща существенная нестационарность.

Все эти проблемы в значительной мере преодолеваются при использовании методологии фильтрации Калмана.

Рассмотрим динамическую систему, чье положение в момент времени t_k описывается линейным векторным разностным уравнением

$$\mathbf{x}_k = \Phi_{k,k-1} \mathbf{x}_{k-1} + \Gamma_{k,k-1} \mathbf{u}_{k-1} + \mathbf{f}_{k-1} + \mathbf{w}_{k-1}, \quad (3.2.41)$$

где $\mathbf{x}_k$ – вектор, размерностью n , описывающий состояние системы (многомерный временной ряд);

$\Phi_{k,k-1}$ – матрица, размером $n \times n$, переходов системы из состояния в момент времени t_{k-1} в состояние, соответствующее моменту времени t_k ;

$\mathbf{u}_{k-1}$ – вектор контроля за системой на полуинтервале времени $[t_{k-1}, t_k)$, размерностью p ;

$\Gamma_{k,k-1}$ – известная матрица, размером $n \times p$, связанная с контролем состояния системы, благодаря которому вектор контроля изменяется от одного момента времени к другому;

$\mathbf{f}_{k-1}$ – вектор, порядка p , силовой функции, представляющей собой известную функцию времени;

$\mathbf{w}_{k-1}$ – вектор, порядка p , случайной последовательности (вектор случайных отклонений) с известными свойствами:

$$\mathcal{M}\{\mathbf{w}_k\} = \mathbf{0}, \text{ для всех } k;$$

$$\mathcal{M}\{\mathbf{w}_k \mathbf{w}_j^T\} = \delta_{kj} \mathbf{Q}_k,$$

где δ_{kj} – дельта Кронекера $\left(\delta_{kj} = \begin{cases} 1, & \text{если } k = j \\ 0, & \text{если } k \neq j \end{cases}\right)$; $\mathbf{Q}_k$ – неотрицательно определенная матрица ковариаций между случайными отклонениями в момент времени t_k .

Матрица переходов $\Phi_{k,k-1}$ удовлетворяет следующим свойствам:

$$\begin{aligned} \Phi_{k,k} &= \mathbf{I}, \text{ для всех } k \\ \Phi_{k,j} \Phi_{j,i} &= \Phi_{k,i}. \end{aligned} \quad (3.2.42)$$

Данные свойства приводят к тому, что

$$\Phi_{k,j}^T = \Phi_{j,k}. \quad (3.2.43)$$

Важным моментом в модели фильтра Калмана является начальное состояние системы $\mathbf{x}_0$, которое рассматривается как вектор случайной переменной с известными свойствами

$$\mathcal{M}\{\mathbf{x}_0\} = \mathbf{0}, \quad \mathcal{M}\{\mathbf{x}_0 \mathbf{x}_0^T\} = \mathbf{M}_0 \quad (3.2.44)$$

и

$$\mathcal{M}\{\mathbf{w}_k \mathbf{x}_0^T\} = 0 \quad \text{для всех } k. \quad (3.2.45)$$

Далее предполагается, что в каждый момент времени t_k имеется m измерений, составляющих вектор $\mathbf{z}_k$, который линейно связан с вектором состояния системы и искажен аддитивным шумом:

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad (3.2.46)$$

где $\mathbf{H}_k$ – известная матрица измерений, размером $m \times n$;

$\mathbf{v}_k$ – вектор, порядка m , аддитивной случайной последовательности, обладающий следующими свойствами

$$\begin{aligned} \mathcal{M}\{\mathbf{v}_k\} &= \mathbf{0}, \quad \text{для всех } k; \\ \mathcal{M}\{\mathbf{v}_k \mathbf{v}_j^T\} &= \delta_{kj} \mathbf{R}_k, \end{aligned} \quad (3.2.47)$$

где $\mathbf{R}_k$ – неотрицательно определенная матрица ковариации аддитивного шума в момент времени t_k .

Далее предполагается, что случайные последовательности $\mathbf{v}_k$ и $\mathbf{w}_k$ – некоррелированы. Эти последовательности являются последовательностями «белого шума» с известным свойством:

$$\mathcal{M}\{\mathbf{v}_k \mathbf{w}_j^T\} = \mathbf{0}, \quad \text{для всех } k, j. \quad (3.2.48)$$

Кроме того, случайная последовательность из уравнения (3.2.46) также не коррелирует с вектором начального состояния:

$$\mathcal{M}\{\mathbf{v}_k \mathbf{x}_0^T\} = \mathbf{0}, \quad \text{для всех } k. \quad (3.2.49)$$

После того, как сконструирована модель фильтра Калмана, а она включает два уравнения: (3.2.41) и (3.2.46), а также все сформулированные выше предположения относительно основных свойств параметров и случайных величин, можно поставить задачу рекурсивного линейного оценивания.

На основании модели фильтра Калмана определить оценку $\hat{\mathbf{x}}_k$ состояния системы в момент времени t_k такую, что она является линейной комбинацией оценки состояния системы в момент времени $\hat{\mathbf{x}}_{k-1}$ и измерения данных $\mathbf{z}_k$. Оценка должна быть «лучшей» в смысле того, что ожидаемая величина суммы квадратов ошибок при оценивании является минимальной. Таким образом, оценка $\hat{\mathbf{x}}_k$ должна быть выбрана так, чтобы

$$\mathcal{M}\{(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T (\hat{\mathbf{x}}_k - \mathbf{x}_k)\} = \text{minimum}$$

Калман использовал свою модель, главным образом, для решения проблем управления динамическими системами в техническом приложении. Поэтому первое уравнение модели (3.2.41) имеет силовые компоненты, имеющие отношение к управлению системой. Для решения задач исследования систем, в которых отсутствует силовая составляющая, используется более простая модификация модели.

Пусть динамика системы описывается однородным линейным разностным уравнением

$$\mathbf{x}_k = \Phi_{k,k-1} \mathbf{x}_{k-1}. \quad (3.2.50)$$

Непосредственно вектор состояния системы $\mathbf{x}_k$ не наблюдается, но имеется вектор измерений $\mathbf{z}_k$ (3.2.46):

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k.$$

Таким образом, упрощенная модель фильтра Калмана описывается уравнениями (3.2.50) и (3.2.46).

Первое, форма уравнения линейной оценки состояния может быть установлена на основании физических характеристик системы. Состояние системы развивается на основании уравнения (3.2.50) так, что имея оценку состояния системы $\hat{\mathbf{x}}_{k-1}$ в момент времени t_k , можно спрогнозировать оценку для момента времени t_k на основании выражения:

$$\hat{\mathbf{x}}_k' = \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1}, \quad (3.2.51)$$

если никакая другая информация недоступна. Измерение в момент времени t_k может быть использовано для улучшения оценки. Основываясь на

$\hat{\mathbf{x}}_k'$ и (3.2.46), можно было бы ожидать, что изменение величин в момент времени t_k был бы равен $\mathbf{H}_k \hat{\mathbf{x}}_k'$. Ошибка при оценивании непосредственно отражается на ошибке ожидаемой величине изменений:

$$\mathbf{e}_k = \mathbf{z}_k - \mathbf{H}_k \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1}. \quad (3.2.52)$$

В соответствии с поставленной задачей рекурсивного оценивания, оценка вектора состояния системы должна быть линейной относительно новых измерений. Определим неизвестную матрицу $\mathbf{K}_k$ так, чтобы оценка $\hat{\mathbf{x}}_k$ была найдена на основании выражения:

$$\hat{\mathbf{x}}_k = \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1}). \quad (3.2.53)$$

Матрица $\mathbf{K}_k$ должна быть определена так, чтобы значение выражения

$$\mathcal{M}\{(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T (\hat{\mathbf{x}}_k - \mathbf{x}_k)\}$$

было минимальным, при этом она может рассматриваться как весовая матрица.

Обозначим разность между оценкой состояния системы и ее истинным значением как:

$$\tilde{\mathbf{x}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k,$$

таким образом,

$$\mathcal{M}\{(\hat{\mathbf{x}}_k - \mathbf{x}_k)^T (\hat{\mathbf{x}}_k - \mathbf{x}_k)\} = \mathcal{M}\{\tilde{\mathbf{x}}_k^T \tilde{\mathbf{x}}_k\}. \quad (3.2.54)$$

Очевидно, что выражение (3.2.54) может быть переписано следующим образом:

$$\mathcal{M}\{\tilde{\mathbf{x}}_k^T \tilde{\mathbf{x}}_k\} = \text{tr}[\mathcal{M}\{\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T\}],$$

где $\text{tr}[\cdot]$ – след матрицы, т.е. оператор взятия суммы диагональных элементов матрицы, стоящей в квадратных скобках. Определим матрицу $\mathbf{P}_k$ как

$$\mathbf{P}_k = \mathcal{M}\{\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T\}. \quad (3.2.55)$$

Теперь получим выражение для $\tilde{\mathbf{x}}_k$:

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \left(\Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1}) \right) = \\ &= \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} - \mathbf{K}_k \mathbf{H}_k \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k (\mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k) = \\ &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \mathbf{v}_k. \end{aligned} \quad (3.2.56)$$

Выражение (3.2.56) можно подставить в (3.2.55) и получить выражение для матрицы $\mathbf{P}_k$

$$\begin{aligned} \mathbf{P}_k = & \mathcal{M} \left\{ \left((\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \boldsymbol{\Phi}_{k,k-1} \tilde{\mathbf{x}}_{k-1} + \mathbf{K}_k \mathbf{v}_k \right) \left((\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \boldsymbol{\Phi}_{k,k-1} \tilde{\mathbf{x}}_{k-1} + \mathbf{K}_k \mathbf{v}_k \right)^T \right\} = \\ & (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \boldsymbol{\Phi}_{k,k-1} \mathcal{M} \{ \tilde{\mathbf{x}}_{k-1} \tilde{\mathbf{x}}_{k-1}^T \} \boldsymbol{\Phi}_{k,k-1}^T (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \\ & + \mathbf{K}_k \mathcal{M} \{ \mathbf{v}_k \tilde{\mathbf{x}}_{k-1}^T \} \boldsymbol{\Phi}_{k,k-1}^T (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \\ & + (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \boldsymbol{\Phi}_{k,k-1} \mathcal{M} \{ \tilde{\mathbf{x}}_{k-1} \mathbf{v}_k^T \} \mathbf{K}_k^T + \mathbf{K}_k \mathcal{M} \{ \mathbf{v}_k \mathbf{v}_k^T \} \mathbf{K}_k^T. \quad (3.2.57) \end{aligned}$$

По определению и по аналогии с (3.2.55), определим матрицу

$$\mathbf{P}_{k-1} = \mathcal{M} \{ \tilde{\mathbf{x}}_{k-1} \tilde{\mathbf{x}}_{k-1}^T \},$$

а также укажем на матрицу из (3.2.47):

$$\mathcal{M} \{ \mathbf{v}_k \mathbf{v}_k^T \} = \mathbf{R}_k.$$

Рассматривая также следующие свойства введенных величин, а именно

$$\mathcal{M} \{ \mathbf{v}_k \tilde{\mathbf{x}}_{k-1}^T \} = \mathcal{M} \{ \tilde{\mathbf{x}}_{k-1} \mathbf{v}_k^T \} = \mathbf{0},$$

$$\mathcal{M} \{ \mathbf{v}_k \mathbf{v}_{k-1}^T \} = \mathbf{0},$$

и

$$\mathcal{M} \{ \mathbf{v}_k \mathbf{x}_0^T \} = \mathbf{0},$$

выражение (3.2.57) можно переписать в следующем виде:

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}'_k (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T, \quad (3.2.58)$$

где использовано обозначение матрицы

$$\mathbf{P}'_k = \boldsymbol{\Phi}_{k,k-1} \mathbf{P}_{k-1} \boldsymbol{\Phi}_{k,k-1}^T. \quad (3.2.59)$$

Раскрывая скобки в (3.2.58) и после небольших преобразований, получаем

$$\mathbf{P}_k = \mathbf{P}'_k - \mathbf{K}_k \mathbf{H}_k \mathbf{P}'_k - \mathbf{P}'_k \mathbf{H}_k^T \mathbf{K}_k^T + \mathbf{K}_k (\mathbf{H}_k \mathbf{P}'_k \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{K}_k^T. \quad (3.2.60)$$

Матрица $\mathbf{P}'_k$ не зависит от $\mathbf{K}_k$, поэтому на нее не оказывает эффект выбор матрицы $\mathbf{K}_k$. Матрица $(\mathbf{H}_k \mathbf{P}'_k \mathbf{H}_k^T + \mathbf{R}_k)$ является симметричной и неотрицательно-определенной, так что она может быть переписана как результат произведения матрицы $\mathbf{S}_k$ на себя транспонированную, т.е. $\mathbf{S}_k$ — матрица корня квадратного:

$$\mathbf{S}_k \mathbf{S}_k^T = \mathbf{H}_k \mathbf{P}'_k \mathbf{H}_k^T + \mathbf{R}_k. \quad (3.2.61)$$

Заметим, что последние три члена выражения (3.2.60) имеют форму квадратичного матричного полинома относительно неизвестной матрицы $\mathbf{K}_k$. Подставим (3.2.61) в выражение (3.2.60) и предположим существование матрицы $\mathbf{A}_k$ такой, что выполняется тождество

$$\mathbf{P}_k = \mathbf{P}'_k + (\mathbf{K}_k \mathbf{S}_k - \mathbf{A}_k)(\mathbf{K}_k \mathbf{S}_k - \mathbf{A}_k)^T - \mathbf{A}_k \mathbf{A}_k^T. \quad (3.2.62)$$

Эта процедура является матричным эквивалентом извлечения корня из квадратичного полинома. Предполагая, что $\mathbf{S}_k \mathbf{S}_k^T$ – положительно-определенная матрица, а это непосредственно приводит к выражению

$$\mathbf{A}_k = \mathbf{P}'_k \mathbf{H}_k^T (\mathbf{S}_k^{-1})^T. \quad (3.2.63)$$

Когда матрица $\mathbf{S}_k$ – сингулярная, то вместо обратной матрицы используется процедура псевдообращения матрицы. В выражении (3.2.63) только один член – результат произведения – зависит от весовой матрицы $\mathbf{K}_k$. Произведение матрицы на себя транспонированную – матрица неотрицательно-определенная, поэтому след матрицы $\mathbf{P}_k$ минимизируется посредством выбора

$$\mathbf{K}_k \mathbf{S}_k = \mathbf{P}'_k \mathbf{H}_k^T (\mathbf{S}_k^{-1})^T,$$

таким образом, оптимальная весовая матрица должна иметь вид:

$$\mathbf{K}_k = \mathbf{P}'_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}'_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1}. \quad (3.2.64)$$

Подставляя (3.2.64) и (3.2.63) в выражение (3.2.62), получаем окончательное выражение для матрицы $\mathbf{P}_k$:

$$\mathbf{P}_k = \mathbf{P}'_k - \mathbf{K}_k \mathbf{H}_k \mathbf{P}'_k. \quad (3.2.65)$$

Уравнения (3.2.53), (3.2.59), (3.2.64) и (3.2.65) составляют фильтр Калмана для модели (3.2.50) и (3.2.46).

Расширим модель, включив в нее детерминистическую и случайную силовые составляющие:

$$\mathbf{x}_k = \Phi_{k,k-1} \mathbf{x}_{k-1} + \mathbf{f}_{k-1} + \mathbf{w}_{k-1}, \quad (3.2.66)$$

в этой модели отсутствует компонента контроля.

Предположим, что оценка состояния системы основана только на оценке в момент времени t_{k-1} , т.е. никаких измерений в момент времени t_k не

производится. Вектор шума $\mathbf{w}_k$ не зависит от состояния в момент t_k и нулевое математическое ожидание, таким образом вполне ожидаемо, что он не оказывает влияние на оценку состояния в момент времени t_k . Также, $\mathbf{f}_{k-1}$ – известная векторная функция, которая действует на полуинтервале $[t_{k-1}, t_k)$, поэтому оценка может быть получена на основании выражения

$$\hat{\mathbf{x}}'_k = \Phi_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{f}_{k-1}. \quad (3.2.67)$$

Покажем, что эта оценка является наилучшей с точки зрения критерия минимума наименьших квадратов.

В теории метода наименьших квадратов известно утверждение, согласно которому наилучшей оценкой случайного вектора $\mathbf{x}$ на основании известных величин $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q$, если известна плотность их совместного распределения $p(\mathbf{x}, \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q)$, является условное математическое ожидание случайной величины $\mathbf{x}$ при наличии заданных значений $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_q$:

$$\hat{\mathbf{x}} = \mathcal{M}\{\mathbf{x} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\}.$$

Таким образом, наилучшей оценкой неизвестного вектора состояний $\mathbf{x}_k$ на основании измерений $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}$ является условное математическое ожидание

$$\begin{aligned} \hat{\mathbf{x}}'_k &= \mathcal{M}\{\mathbf{x}_k | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} = \\ &= \mathcal{M}\{(\Phi_{k,k-1} \mathbf{x}_{k-1} + \mathbf{f}_{k-1} + \mathbf{w}_{k-1}) | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} = \\ &= \Phi_{k,k-1} \mathcal{M}\{\mathbf{x}_{k-1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} + \mathcal{M}\{\mathbf{f}_{k-1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} + \\ &\quad + \mathcal{M}\{\mathbf{w}_{k-1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\}. \end{aligned} \quad (3.2.68)$$

Первое математическое ожидание в (3.2.68) может быть найдено по определению:

$$\mathcal{M}\{\mathbf{x}_{k-1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} = \hat{\mathbf{x}}_{k-1}. \quad (3.2.69)$$

$\mathbf{f}_{k-1}$ – не случайный вектор, поэтому

$$\mathcal{M}\{\mathbf{f}_{k-1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} = \mathbf{f}_{k-1}. \quad (3.2.70)$$

В основных предположениях модели фильтра Калмана отмечалось, что вектор шума $\mathbf{w}_{k-1}$ не зависит от состояний системы для всех моментов времени, и также не зависит от погрешностей измерений, поэтому:

$$\mathcal{M}\{\mathbf{w}_{k-1}|\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}\} = \mathbf{0}. \quad (3.2.71)$$

Подставляя выражения (3.2.69), (3.2.70) и (3.2.71) в уравнение (3.2.68), получаем наилучшую оценку состояния системы, основанной на оценке в момент t_{k-1} :

$$\hat{\mathbf{x}}'_k = \Phi_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{f}_{k-1},$$

таким образом, получена оценка (3.2.67).

Предположим, что произведено измерение в момент времени t_k . Аналогично предположениям, которые были произведены для оценки (3.2.53), уточненную оценку определим следующим образом:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}'_k + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}'_k), \quad (3.2.72)$$

где $\mathbf{K}_k$ — неизвестная весовая матрица. Рассмотрим матрицу $\tilde{\mathbf{x}}_k$

$$\begin{aligned} \tilde{\mathbf{x}}_k &= \hat{\mathbf{x}}_k - \mathbf{x}_k = \left(\Phi_{k,k-1}\hat{\mathbf{x}}_{k-1} + \mathbf{f}_{k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}'_k) \right) - \\ &\quad - \Phi_{k,k-1}\mathbf{x}_{k-1} - \mathbf{f}_{k-1} - \mathbf{w}_{k-1} = \\ &= \Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{K}_k\mathbf{H}_k\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} + \mathbf{K}_k\mathbf{H}_k\mathbf{w}_{k-1} - \mathbf{w}_{k-1} + \mathbf{K}_k\mathbf{v}_k = \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{w}_{k-1}) + \mathbf{K}_k\mathbf{v}_k. \end{aligned} \quad (3.2.73)$$

Рассмотрим матрицу:

$$\begin{aligned} \mathbf{P}_k &= \mathcal{M}\left\{ \left((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{w}_{k-1}) + \mathbf{K}_k\mathbf{v}_k \right) \cdot \right. \\ &\quad \left. \cdot \left((\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{w}_{k-1}) + \mathbf{K}_k\mathbf{v}_k \right)^T \right\} \end{aligned}$$

Поскольку математическое ожидание квадратичной формы имеет вид:

$$\begin{aligned} \mathcal{M}\left\{ (\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{w}_{k-1})(\Phi_{k,k-1}\tilde{\mathbf{x}}_{k-1} - \mathbf{w}_{k-1})^T \right\} &= \\ &= \Phi_{k,k-1}\mathbf{P}_{k-1}\Phi_{k,k-1}^T + \mathbf{Q}_{k-1}, \end{aligned}$$

на основании статистических свойств вектора шума $\mathbf{w}_{k-1}$.

Определим матрицу $\mathbf{P}'_k$ следующим образом:

$$\mathbf{P}'_k = \Phi_{k,k-1}\mathbf{P}_{k-1}\Phi_{k,k-1}^T + \mathbf{Q}_{k-1}. \quad (3.2.74)$$

Таким образом, матрица $\mathbf{P}_k$ имеет следующий вид:

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}'_k(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T, \quad (3.2.75)$$

и форма уравнения (3.2.75) ничем не отличается от уравнения (3.2.58), единственное отличие состоит в значении матрицы P'_k . Следовательно, оптимальная весовая матрица K_k получается на основании выражения (3.2.64), а матрица P_k в конечном итоге может быть найдена по выражению (3.2.65).

Уравнения (3.2.72), (3.2.74), (3.2.64) и (3.2.65) устанавливают фильтр Калмана для модели с силовыми компонентами, принимающей вид уравнений (3.2.66) и (3.2.46). Следует отметить, что единственный эффект случайного шума w_{k-1} , действующего в динамической системе, проявляется в увеличении неопределенности при оценивании благодаря аддитивному присутствию в выражении (3.2.74). С другой стороны, детерминистическая составляющая f_k появляется лишь в уравнении оценки (3.2.72). Во всем остальном модель фильтра с силовыми компонентами (3.2.66) ничем не отличается от модели, представленной линейным однородным разностным уравнением (3.2.50). Поскольку алгоритмы управления, используемые в общей модели фильтра Калмана (3.2.41), предназначены для технических систем, что с большим трудом может быть применено к экономическим системам, то проблемы оценивания параметром модели с компонентами управления рассматриваться не будут. С решением проблем оценивания при наличии в модели компонент управления можно познакомиться в работе автора [35] и обзорной работе, посвященной всем различным аспектам фильтра Калмана [36].

3.3. Факторный подход к модели многомерных временных рядов

В анализе многомерных временных рядов возникает много проблем [28, с.145], некоторые, и в частности, проблема идентифицируемости к настоящему моменту времени не решены до конца. Как отметил М. Кендэл, «проблема идентифицируемости все еще ждет полного и удобного для практики решения» [28, с. 146]. Более того, механистическое обобщение одномерного случая на многомерные временные ряды, которое сводится к тому, что рассматриваются

три вида моделей: авторегрессии, скользящего среднего и смешанные, – ведет к нивелированию проблемы разделения временного влияния на значение переменной в данный момент наблюдения и корреляции с другими переменными.

Традиционные подходы анализа многомерных временных рядов предполагают влияние на данную переменную в данный момент времени как ее значений в предшествующие моменты, так и значений других переменных, и в текущий момент, и в предшествующие. Для этого вводится понятие кросс-корреляции между различными переменными, причем разработанный аппарат позволяет рассматривать характеристики тесноты связи среди переменных лишь между их парами. У спектральной плотности кросс-спектра для двух рядов наряду с действительной частью появляется и мнимая часть, и в этой связи возникают еще две характеристики: когерентность – отношение амплитуды кросс-спектра к произведению спектральных плотностей каждого из двух рядов, – и, так называемое, «усиление» – произведение отношения спектральных плотностей двух рядов на когерентность.

«Интерпретация этих величин в анализе временных рядов, как мне кажется, дело более трудное, чем интерпретация кросс-коррелограммы, которая, по крайней мере, показывает силу связи между двумя рядами при различных опережениях или запаздываниях», – отмечал М. Кендэл [28]. Таким образом, слабая интерпретабельность получаемых результатов, или ее полная невозможность, глубокая запутанность воздействий между переменными для разных моментов времени делают анализ многомерных временных рядов малопривлекательным, и в этой связи – и малораспространенным.

Главной нерешенной проблемой анализа многомерных временных рядов является невозможность отдельного учета влияния временных сдвигов и влияния других переменных. Ответ на вопрос, что привело к данному значению переменной: либо ее значение в предыдущие моменты времени, либо влияние других переменных, – является центральным в многомерном анализе временных рядов как с точки зрения выяснения характера поведения

экономической системы, так и для корректного решения проблем прогнозирования. Поэтому в самодостаточной модели многомерного временного ряда должно быть разделение временного аспекта и проблемы взаимодействий между различными переменными, между которыми существует объективная связь.

Пусть имеется несколько переменных, наблюдения для каждой из которых образуют временной ряд. Естественно, что наблюдения произведены в одни и те же моменты времени и их число значительно больше числа переменных. Для построения модели сделаем основное предположение. Значение величины переменной в момент времени t непосредственно не зависит от ее значений в моменты $t - 1, t - 2, \dots$. Но поведение совокупности переменных, как и в обычном факторном анализе, определяется поведением факторов. Именно факторы подвержены влиянию временных сдвигов. Величина отдельного значения фактора для данного момента времени зависит только от того, какие значения фактор принимал в предшествующие моменты, и не зависит от значений других факторов, поскольку в ортогональной факторной модели факторы являются независимыми друг от друга переменными. Именно поведение каждого фактора можно рассматривать как автокорреляционный процесс.

Построим модель многомерного временного ряда с разделением проблем, связанных с временем и с корреляцией между исследуемыми признаками.

Будем рассматривать значения наблюдаемых переменных, составляющих многомерный временной ряд в виде вектора наблюдений в момент времени t :

$$\mathbf{z}_t = (z_{t1} \quad z_{t2} \quad \dots \quad z_{tn}),$$

где n – число переменных. Также как в обычном факторном анализе будем рассматривать наблюдаемые значения переменных в стандартной форме, т.е. они должны быть центрированными и нормированными. Если речь идет о временных рядах, то строго говоря, у наблюдаемой переменной должен быть удален тренд, который может иметь и нелинейный характер. Однако на этом

этапе исследования будем полагать, что вероятностные характеристики наблюдаемой переменной полностью описываются математическим ожиданием и дисперсией в силу основного предположения, что эта переменная непосредственно не зависит от времени.

Обозначим вектор значений факторов, определяющих поведение наблюдаемых переменных в момент времени t , как

$$\mathbf{f}_t = (f_{t1} \ f_{t2} \ \dots \ f_{tm}),$$

где m – число факторов.

Для получения обобщенной многомерной авторегрессионной модели факторов, рассмотрим сначала случай авторегрессии первого порядка:

$$\mathbf{f}_t = \mathbf{f}_{t-1}\Phi_1 + \mathbf{a}_t, \quad (3.3.1)$$

где Φ_1 – матрица перехода значений факторов от момента времени $(t - 1)$ к t ,

$\mathbf{a}_t$ – вектор «белого шума», или импульсов, или случайных отклонений

$$\mathbf{a}_t = (a_{t1} \ a_{t2} \ \dots \ a_{tm}).$$

Пусть, как и для одномерного случая, B – оператор сдвига назад, тогда выражение (3.3.1) можно записать в следующем виде

$$\mathbf{f}_t - \mathbf{f}_{t-1}\Phi_1 = \mathbf{a}_t$$

или

$$\mathbf{f}_t - B\mathbf{f}_t\Phi_1 = \mathbf{a}_t,$$

и тогда модель (3.3.1) можно представить следующим образом:

$$\mathbf{f}_t(I - B\Phi_1) = \mathbf{a}_t. \quad (3.3.2)$$

В этой модели неизвестными параметрами, подлежащими оценке, являются матрица перехода на один шаг времени вперед Φ_1 , а также дисперсии m компонент вектора белого шума.

Если предположить, что переход значений факторов от момента времени t к моменту времени $t - 2$ определяется произведением матриц перехода, что логично следует из следующего соотношения:

$$\mathbf{f}_{t-1} = \mathbf{f}_{t-2}\Phi_1 + \mathbf{a}_{t-1}, \quad (3.3.3)$$

где, подставляя (3.3.3) в (3.3.1)

$$f_t = (f_{t-2}\Phi_1 + a_{t-1})\Phi_1 + a_t,$$

получаем

$$f_t = f_{t-2}\Phi_1\Phi_1 + a_{t-1}\Phi_1 + a_t = f_{t-2}\Phi_1^2 + a_{t-1}\Phi_1 + a_t, \quad (3.3.4)$$

то модель авторегрессии факторов 2-го порядка имеет вид

$$f_t = f_{t-1}\Phi_1 + f_{t-2}\Phi_1^2 + a_t,$$

или, используя оператор сдвига назад:

$$f_t(I - B\Phi_1 - B^2\Phi_1^2) = a_t. \quad (3.3.5)$$

Процедура, использованная при получении (3.3.4), если ее применить несколько раз, ведет к модели фильтра. В модели авторегрессии, в отличие от модели фильтра, вместо значений шума в предыдущие моменты времени используются значения факторов. В модели авторегрессии 2-го порядка, также как в модели первого порядка, такое же число неизвестных параметров.

Обобщая (3.3.5) на случай авторегрессии произвольного порядка, получаем модель следующего вида

$$f_t(I - \sum_{j=1}^p B^j\Phi_1^j) = a_t, \quad (3.3.6)$$

где p – порядок авторегрессии. Поскольку матрица перехода из одного момента времени в любой другой выражается через матрицу перехода между соседними моментами времени, то далее ее будем обозначать без индекса, т.е. просто – Φ .

Таким образом, многомерная авторегрессионная модель факторов порядка p имеет следующий вид:

$$f_t(I - \sum_{j=1}^p B^j\Phi) = a_t. \quad (3.3.7)$$

Значения исследуемых переменных в текущий момент времени определяются значениями факторов в текущий момент:

$$z_t = f_t A^T + v_t, \quad (3.3.8)$$

где A – матрица факторных нагрузок, значения которых не меняются с течением времени;

v_t – значения произведения характерных факторов на соответствующие нагрузки в момент времени t , рассматриваются как погрешности факторного объяснения наблюдаемых переменных.

Рассмотрим модель факторного анализа временных рядов с авторегрессией первого порядка факторов. Она состоит из двух уравнений: уравнения авторегрессии факторов (3.3.1) и уравнения зависимости наблюдаемых переменных от факторов (3.3.8), т.е. определяется системой уравнений

$$\begin{cases} f_t = f_{t-1}\Phi + a_t \\ z_t = f_t A^T + v_t \end{cases} \quad (3.3.9)$$

Сделаем для нее основные предположения.

Математическое ожидание случайных отклонений факторной авторегрессии равно нулю для всех моментов времени:

$$\mathcal{M}\{a_t\} = 0, t = 1, 2, \dots \quad (3.3.10)$$

Случайные отклонения факторной авторегрессии имеют постоянную дисперсию σ^2 для любого момента времени:

$$\mathcal{M}\{a_t^T a_t\} = \sigma^2 I, t = 1, 2, \dots \quad (3.3.11)$$

где I – единичная матрица порядка m .

Случайные отклонения уравнения факторной авторегрессии для разных моментов времени не связаны друг с другом, т.е. между ними отсутствует корреляция:

$$\mathcal{M}\{a_t^T a_{t-s}\} = 0, t, s = 1, 2, \dots, t \neq s. \quad (3.3.12)$$

Поскольку наблюдаемые величины центрированы и нормированы, т.е. приведены к стандартному виду, то этому виду должны соответствовать и значения факторов. Поэтому очередное предположение следующее: математическое ожидание значений факторов для всех моментов времени равно нулю:

$$\mathcal{M}\{f_t\} = 0, t = 0, 1, 2, \dots \quad (3.3.13)$$

Значения факторов для начального момента времени – ортогональны:

$$f_0^T f_0 = I. \quad (3.3.14)$$

Факторы и случайные отклонения не коррелируют друг с другом:

$$\mathcal{M}\{\mathbf{f}_t^T \mathbf{a}_t\} = \mathbf{0}, t = 1, 2, \dots \quad (3.3.15)$$

Матрица перехода Φ должна иметь общее для подобного рода матриц свойство, а именно: чтобы вычислить по отношению к Φ обратную матрицу, достаточно ее транспонировать:

$$\Phi^T = \Phi^{-1}. \quad (3.3.16)$$

Этому свойству удовлетворяют матрицы, составленные из собственных векторов некоторой симметричной матрицы.

Погрешности уравнения зависимости наблюдаемых переменных от факторов имеют нулевое математическое ожидание:

$$\mathcal{M}\{\mathbf{v}_t\} = \mathbf{0}, t = 1, 2, \dots \quad (3.3.17)$$

Эти погрешности не коррелируют ни с наблюдаемыми переменными, ни с факторами, ни со случайными отклонениями факторной авторегрессии:

$$\mathcal{M}\{\mathbf{z}_t^T \mathbf{v}_t\} = \mathcal{M}\{\mathbf{f}_t^T \mathbf{v}_t\} = \mathcal{M}\{\mathbf{a}_t^T \mathbf{v}_t\} = \mathbf{0}, t = 1, 2, \dots \quad (3.3.18)$$

Погрешности второго уравнения системы (3.3.9) имеют постоянный с течением времени момент второго порядка:

$$\mathcal{M}\{\mathbf{v}_t^T \mathbf{v}_t\} = \mathbf{D}^2, t = 1, 2, \dots, \quad (3.3.19)$$

где $\mathbf{D}^2$ – диагональная матрица дисперсий погрешностей, которая по аналогии с традиционным факторным анализом может быть названа матрицей значений характеристик.

Погрешности $\mathbf{v}$ для разных моментов времени не коррелируют между собой:

$$\mathcal{M}\{\mathbf{v}_t^T \mathbf{v}_{t-s}\} = \mathbf{0}, t, s = 1, 2, \dots, t \neq s. \quad (3.3.20)$$

Уравнение (3.3.1) предполагает, что значения факторов в некоторый момент времени являются случайными величинами. Рассмотрим некоторые свойства этих случайных величин, следующие из принятых предположений. Для этого преобразуем уравнение авторегрессии (3.3.1) в модель фильтра:

$$\begin{aligned} \mathbf{f}_t &= (\mathbf{f}_{t-2} \Phi + \mathbf{a}_{t-1}) \Phi + \mathbf{a}_t = \dots = \mathbf{f}_0 \Phi^t + \mathbf{a}_1 \Phi^{t-1} + \\ &+ \mathbf{a}_2 \Phi^{t-2} + \dots + \mathbf{a}_{t-1} \Phi + \mathbf{a}_t = \mathbf{f}_0 \Phi^t + \sum_{i=1}^t \mathbf{a}_i \Phi^{t-i}. \end{aligned} \quad (3.3.21)$$

Используя выражение (3.3.21) и предположение о свойствах случайных величин модели (3.3.9), определим ковариацию значений факторов для момента времени t . Для этого найдем математическое ожидание произведения значений факторов:

$$\begin{aligned}\mathcal{M}\{f_t^T f_t\} &= (f_0 \Phi^t + \sum_{i=1}^t a_i \Phi^{t-i})^T (f_0 \Phi^t + \sum_{i=1}^t a_i \Phi^{t-i}) = \\ &= I + \sum_{i=1}^t \sigma^2 I = I + t\sigma^2 I = (1 + t\sigma^2)I.\end{aligned}\quad (3.3.22)$$

Поскольку матрица ковариаций – диагональная, то значения факторов для любого момента времени – ортогональны, но на диагонали этой матрицы находятся значения большие единицы.

Матрица ковариаций значений факторов для моментов времени, расположенных рядом, имеет вид

$$\mathcal{M}\{f_{t-1}^T f_t\} = \mathcal{M}\{f_{t-1}^T f_{t-1} \Phi + f_{t-1}^T a_t\} = (1 + (t-1)\sigma^2)\Phi. \quad (3.3.23)$$

Выражение (3.3.23) показывает, что эта матрица уже не диагональная, поэтому свойство ортогональности значений факторов для разных моментов времени не выполняется, и ковариации между факторами определяются матрицей перехода. В частности, для произвольных моментов времени матрица ковариаций между значениями факторов определяется следующим выражением:

$$\begin{aligned}\mathcal{M}\{f_{t-s}^T f_t\} &= \mathcal{M}\left\{(f_0 \Phi^{t-s} + \sum_{i=1}^{t-s} a_i \Phi^{t-s-i})^T (f_0 \Phi^t + \sum_{i=1}^t a_i \Phi^{t-i})\right\} = \\ &= \Phi^s + \sum_{i=1}^{t-s} (\Phi^{t-s-i})^T a_i^T a_i \Phi^{t-s-i} = \Phi^s + (t-s)\sigma^2 \Phi^s = \\ &= (1 + (t-s)\sigma^2)\Phi^s.\end{aligned}\quad (3.3.24)$$

И в частности, когда $t = s$, имеем

$$\mathcal{M}\{f_0^T f_t\} = \Phi^t, \quad (3.3.25)$$

т.е. ковариация значений факторов в некоторый момент времени со значениями в начальный момент времени не зависит от дисперсии случайных отклонений авторегрессии факторов.

Теперь рассмотрим связь между факторами и значениями наблюдаемых переменных. Естественно, что эта связь определяется уравнением (3.3.8),

поэтому рассмотрим ковариацию между факторами и исследуемыми признаками

$$\mathcal{M}\{\mathbf{f}_t^T \mathbf{z}_t\} = \mathcal{M}\{\mathbf{f}_t^T (\mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t)\} = \mathcal{M}\{\mathbf{f}_t^T \mathbf{f}_t \mathbf{A}^T + \mathbf{f}_t^T \mathbf{v}_t\} = (1 + t\sigma^2) \mathbf{A}^T. \quad (3.3.26)$$

Таким образом, ковариация между факторами и переменными определяется так же, как и в обычном факторном анализе, матрицей факторных нагрузок, но скорректированной на временной множитель $(1 + t\sigma^2)$. Однако для начального момента времени выполняется одна из теорем классического факторного анализа, а именно:

$$\mathcal{M}\{\mathbf{f}_0^T \mathbf{z}_0\} = \mathbf{A}^T. \quad (3.3.27)$$

Ковариация между факторами для некоторого момента времени и значениями наблюдаемых переменных для другого момента времени, число шагов между которыми равно s можно определить на основании уравнений (3.3.26) и (3.3.24), и она выглядит следующим образом

$$\mathcal{M}\{\mathbf{f}_t^T \mathbf{z}_{t-s}\} = (1 + (t-s)\sigma^2) \Phi^s \mathbf{A}^T, \quad (3.3.28)$$

Рассмотрим вопрос о корреляции между наблюдаемыми переменными для некоторого момента времени t , которые согласно догмам факторного анализа объясняются исключительно факторными нагрузками. Для этого воспользуемся моделью (3.3.9). Матрица корреляций между переменными имеет вид:

$$\begin{aligned} \mathbf{R} &= \mathcal{M}\{\mathbf{z}_t^T \mathbf{z}_t\} = \mathcal{M}\{(\mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t)^T (\mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t)\} = \\ &= \mathcal{M}\{\mathbf{A} \mathbf{f}_t^T \mathbf{f}_t \mathbf{A}^T + \mathbf{A} \mathbf{f}_t^T \mathbf{v}_t + \mathbf{v}_t^T \mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t^T \mathbf{v}_t\} = \\ &= \mathcal{M}\{\mathbf{A} \mathbf{f}_t^T \mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t^T \mathbf{v}_t\} = (1 + t\sigma^2) \mathbf{A} \mathbf{A}^T + \mathbf{D}^2. \end{aligned} \quad (3.3.29)$$

В выражении (3.3.28) значения на главной диагонали корреляционной матрицы могут и не иметь величину, равную единице из-за наличия временного множителя $(1 + t\sigma^2)$, большего единицы. На самом деле этот вопрос требует дополнительного теоретического исследования. Но, если элементы главной диагонали матрицы $(1 + t\sigma^2) \mathbf{A} \mathbf{A}^T$ меньше единицы, то значения характеристик $\mathbf{D}^2$ будут дополнять эти элементы до единицы, и тогда

выражение (3.3.28) совпадает с основной теоремой факторного анализа. Основным условием для этого является выполнение неравенств

$$aa_{ii} < \frac{1}{1+t\sigma^2}, i = 1, 2, \dots, n, \quad (3.3.30)$$

где aa_{ii} – диагональный элемент матрицы AA^T ,
 n – число наблюдаемых переменных.

Рассмотрим вид основных свойств факторной модели авторегрессии первого порядка факторов в предельном значении, т.е. когда $t \rightarrow \infty$. При исследовании асимптотических свойств оценок параметров статистических моделей в этом случае традиционно предполагается, что значение дисперсии погрешностей стремится к нулю, т.е.

$$\lim_{t \rightarrow \infty} \sigma^2 = 0.$$

Тогда основные тождества факторной модели авторегрессии первого порядка приобретают следующий смысл:

значения факторов в предельном значении не только ортогональны, но и нормированы:

$$\mathcal{M}\{f_t^T f_t\} = I, t \rightarrow \infty, \quad (3.3.31)$$

ковариация между соседними значениями факторов совпадает с матрицей переходов:

$$\mathcal{M}\{f_{t-1}^T f_t\} = \Phi, t \rightarrow \infty, \quad (3.3.32)$$

ковариация между значениями факторов, расположенных на расстоянии s шагов друг от друга равна матрице переходов, возведенной в степень s :

$$\mathcal{M}\{f_{t-s}^T f_t\} = \Phi^s, t \rightarrow \infty, \quad (3.3.33)$$

ковариация между значениями факторов и значениями наблюдаемых переменных определяется матрицей факторных нагрузок:

$$\mathcal{M}\{f_t^T z_t\} = A^T, t \rightarrow \infty, \quad (3.3.34)$$

ковариация между значениями факторов для одного момента времени и значениями наблюдаемых переменных для момента времени, отстоящего на s шагов определяется матрицей переходов, возведенной в степень s и матрицей факторных нагрузок

$$\mathcal{M}\{f_t^T z_{t-s}\} = \Phi^s A^T, t \rightarrow \infty, \quad (3.3.35)$$

для предельного случая корреляция между переменными определяется основной теоремой факторного анализа:

$$R = AA^T + D^2, t \rightarrow \infty. \quad (3.3.36)$$

Одним из главных постулатов факторной модели авторегрессии первого порядка факторов является утверждение, состоящее в том, что исследуемые переменные предшествующих моментов времени не оказывают воздействия на текущее значение переменных, но ковариация между ними формально может быть вычислена. Для предельного случая она имеет вид

$$R_{t,t-s} = A\Phi^s A^T + D^2, t \rightarrow \infty, \quad (3.3.37)$$

где $R_{t,t-s}$ — матрица корреляции между наблюдаемыми переменными для моментов времени, расположенных на расстоянии в s шагов.

3.4. Факторная модель в анализе многомерных временных рядов

Рассмотрим экономическую систему, поведение которой описывается рядом показателей. Показатели являются переменными случайными величинами, связанными, или наоборот, не связанными между собой некоторыми неизвестными заранее зависимостями. Поскольку каждый показатель важен для описания системы, то форма зависимости между переменными не будет рассматриваться на этом этапе исследования, целью которого является получение прогноза для всей совокупности показателей.

Сделаем очевидные, но важные предположения для начала исследования: фиксируемые и измеряемые переменные несут полную информацию об экономическом объекте, поскольку будем считать, что регистрируются значения всех возможных характеристик объекта. Избыточность информации, являющаяся вполне логичной в таком случае, не является проблемой, которую необходимо разрешить на данном этапе, и не может привести к искажению выводов.

Существующие методы получения прогнозов основаны на теории Винера-Колмогорова прогнозирования случайных процессов. А эта теория рассматривает процессы, разложимые в ряд Фурье. При этом условие стационарности, или близости к ней, является существенным.

Будем исходить из того, что случайный процесс (управляемый или нет) может быть как стационарным, так и нестационарным. Причем переменные, описывающие экономический процесс, влияют друг на друга, и предсказание поведения отдельного показателя без учета влияния на него других показателей, будет вести к заведомо ложным выводам.

Итак, переменные взаимодействуют между собой, обуславливая тем самым свое будущее поведение. За этим взаимодействием, или взаимным влиянием стоят некие причины – неизвестные, нерегистрируемые и ненаблюдаемые факторы. Переменные в таком случае можно рассматривать как косвенное измерение этих факторов.

Более того, взаимодействие между переменными обусловлено тем, что они отражают разные стороны одного фактора, – и это с одной стороны, а с другой стороны – факторы являются тем скрытым механизмом, который заставляет взаимодействовать или коррелировать разные переменные величины, отражающие различные стороны экономического объекта.

Факторы, или причины, между собой не связаны. И это предположение делается на основании следующего постулата: если бы такая связь имела бы место, то это означало бы, что за взаимодействием таких факторов стоит одна общая причина, объясняющая поведение связанных факторов. И, чтобы установить закономерности динамических изменений исследуемого процесса, достаточно учесть фактор, стоящий над факторами. В факторном анализе такое предположение является основополагающим: факторы – это ортогональные, т.е. независимые надпеременные.

Таким образом, будем исходить из того, что факторы являются независимыми величинами, имеющими неслучайную природу. Т.е. их поведение некоторым образом, но пока неизвестным, предопределено.

Значения факторов определяют состояние экономической системы, и изменение факторов ведет к тому, что система переходит из состояния в состояние с течением времени. Сам переход обусловлен временным фактором, однако, у такого перехода должны быть общие для разных моментов времени закономерности. Если закономерности переходов отсутствуют, то динамика системы имеет характер броуновского движения. Для экономической системы такое движение возможно, однако более общим будет предложение о некотором «разумном» движении, имеющем хоть какую-нибудь закономерность.

Будем фиксировать переходы в определенные моменты времени, рассматривая значения экономических показателей за соответствующий промежуток времени (год, квартал, месяц, etc).

Если закономерности переходов меняются с течением времени, то экономический процесс является нестационарным. Если закономерности переходов остаются неизменными со временем, то экономический процесс – стационарный.

Будем рассматривать фактор, как причину движения экономической системы, абстрагируясь от его места по отношению к самой системе по следующим соображениям. Характер процессов внутри системы может быть обусловлен либо строго внутренними механизмами взаимодействия отдельных ее элементов, либо только внешними «толчками», либо сочетанием тех и других причин. Очевидно, внешние факторы являются причинами внутренних изменений (обратное трудно представить). Поэтому говорить о «внутренних» факторах, не вступая в противоречие с предположением о независимости факторов, будет некорректно. И фактор, имея, безусловно, внешнее происхождение, может оказаться энергетикой структурных внутренних сдвигов, т.е. частью самой экономической системы.

Таким образом, факторы обуславливают поведение экономической системы и, соответственно, определяют значения показателей, с помощью которых система контролируется. Поэтому, чтобы прогнозировать поведение

системы с минимальной ошибкой, следует прогнозировать значения факторов, а значения экономических показателей восстанавливать по оцененным величинам факторов.

Именно динамику факторов, как совокупности независимых друг от друга величин, каждую из которых можно рассматривать как компоненту вектора, следует трактовать как векторный процесс, подпадающий под общую теорию Винера-Колмогорова прогнозирования случайных процессов [34, с.145].

Эта теория рассматривает, в основном, прогноз одномерной величины на основании ее спектральных характеристик. Многомерные величины трактуются как вектор с независимыми координатами, наличие корреляции между которыми приведет к искажению не только результатов прогноза, - но и к искажению идентификации самого процесса.

Математическая модель

Рассмотрим дискретный вариант изменения в экономической системе: факторы меняют свои значения в определенные моменты времени $t = 0, 1, 2, \dots$

Динамика изменения значений факторов описывается уравнением переходов

$$\mathbf{f}_t = \mathbf{f}_{t-1} \Phi_{t-1,t}, \quad (3.4.1)$$

где $\mathbf{f}_t = (f_{t1}, f_{t2}, \dots, f_{tm})$ – вектор значений факторов в момент времени t , m – число факторов;

$\Phi_{t-1,t}$ – матрица переходов значений факторов от момента времени $t-1$ к моменту t , которая имеет вид:

$$\Phi_{t-1,t} = \begin{pmatrix} \varphi_{11} & \cdots & \varphi_{1m} \\ \vdots & \ddots & \vdots \\ \varphi_{m1} & \cdots & \varphi_{mm} \end{pmatrix}$$

Значения изменяемых переменных задаются уравнением факторной структуры

$$\mathbf{z}_t = \mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t, \quad (3.4.2)$$

где $\mathbf{z}_t = (z_{t1}, z_{t2}, \dots, z_{tm})$ – вектор стандартизированных значений переменных – экономических показателей в момент времени t ; n – число изменяемых показателей; $(\cdot)^T$ знак транспонирования матрицы;

$\mathbf{A}$ – матрица факторных нагрузок, имеющая вид

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

$\mathbf{v}_t = (v_{t1}, v_{t2}, \dots, v_{tm})$ – вектор случайных возмущений уравнения факторной структуры.

Стандартизированное значение экономического показателя означает, что из его реального, наблюдаемого значения вычли среднее значение и эту разность разделили на среднеквадратическое отклонение:

$$z_{tk} = \frac{x_{tk} - \bar{x}_k}{\sigma_k},$$

где x_{tk} – зафиксированное, реальное значение экономического показателя;

$\bar{x}_k$ – среднее значение k -го показателя;

σ_k – его среднеквадратическое отклонение.

По своей математической сути уравнения (3.4.1) и (3.4.2) представляют собой фильтр Калмана, причем (3.4.1) – уравнение переходов, (3.4.2) – уравнение измерений. Однако Калман предполагал, что все матрицы известны и необходимо было оценить значение переданной, но неизвестной информации $\mathbf{f}_t$ по принятому с наложением шумов сигналу $\mathbf{z}_t$. В нашем случае, ни одна матрица, кроме значений наблюдаемых переменных $\mathbf{z}_t$, не известна, неизвестными являются также значения факторов. Предстоит оценить матрицу переходов, матрицу факторных нагрузок и найти значения факторов. Такой постановки задачи у Калмана нет, и соответственно воспользоваться ходом его решения данной задачи не представляется возможным. Поэтому предварительно исследуем свойства случайных отклонений, а также факторов.

Случайные отклонения уравнения переходов и уравнения факторной структуры должны удовлетворять следующим условиям:

- математическое ожидание случайного отклонения равно нулю

$$\mathcal{M}\{\mathbf{v}_t\} = 0, (t = 0, 1, 2, \dots, N),$$

здесь $\mathcal{M}\{\cdot\}$ – символ математического ожидания, и N – число всех наблюдений;

- случайные отклонения не коррелируют с переменными и факторами

$$\begin{aligned}\mathcal{M}\{\mathbf{v}_t^T \mathbf{f}_t\} &= 0; \\ \mathcal{M}\{\mathbf{v}_t^T \mathbf{f}_{t-1}\} &= 0; \quad (t = 1, 2, \dots, N);\end{aligned}$$

- момент второго порядка случайного отклонения уравнения факторной структуры для каждого момента времени равен

$$\mathcal{M}\{\mathbf{v}_t^T \mathbf{v}_t\} = \mathbf{D}^2,$$

где $\mathbf{D}^2$ – диагональная матрица, называемая в факторном анализе матрицей значений характеристик; она представляет собой матрицу дисперсий, так называемых, характерных факторов.

- момент второго порядка случайных отклонений уравнения факторной структуры для разных моментов времени равен нулю:

$$\mathcal{M}\{\mathbf{v}_{t-1}^T \mathbf{v}_t\} = 0$$

Теперь отметим основное предположение для факторов:

факторы в данный момент времени являются ортогональными, и это справедливо для всех моментов времени

$$\mathbf{f}_t^T \mathbf{f}_t = \mathbf{I}; (t = 0, 1, 2, \dots, N),$$

здесь $\mathbf{I}$ – единичная матрица размером $m \times m$.

Следующие утверждения вытекают из выше сделанных предположений.

1) Парные корреляции между переменными образуют матрицу $\mathbf{R}$, которая для всех моментов времени имеет вид

$$\begin{aligned}\mathbf{R} &= \mathcal{M}\{\mathbf{z}_t^T \mathbf{z}_t\} = \mathcal{M}\{(\mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t)^T (\mathbf{f}_t \mathbf{A}^T + \mathbf{v}_t)\} = \\ &= \mathbf{A} \mathbf{f}_t^T \mathbf{f}_t \mathbf{A}^T + \mathbf{A} \mathbf{f}_t^T \mathcal{M}\{\mathbf{v}_t\} + \mathcal{M}\{\mathbf{v}_t^T\} \mathbf{f}_t \mathbf{A}^T + \mathcal{M}\{\mathbf{v}_t^T \mathbf{v}_t\} = \\ &= \mathbf{A} \mathbf{A}^T + \mathbf{D}^2.\end{aligned}$$

Уравнение

$$\mathbf{R} = \mathbf{A} \mathbf{A}^T + \mathbf{D}^2$$

называется «фундаментальной теоремой факторного анализа».

2) Корреляция между переменными и факторами имеет вид:

$$\mathcal{M}\{f_t^T z_t\} = \mathcal{M}\{f_t^T f_t A^T + f_t^T v_t\} = A^T,$$

здесь учтена ортогональность факторов и равенство нулю математического ожидания случайных отклонений. Таким образом, корреляции между переменными и факторами определяют матрицу факторных нагрузок.

3) Переходы значений факторов определяют переходы значений экономических показателей.

Покажем это. Обозначим матрицу переходов значений показателей как $M_{t-1,t}$, т.е.

$$M_{t-1,t} = \mathcal{M}\{z_{t-1}^T z_t\}$$

тогда, поскольку

$$z_{t-1} = f_{t-1} A^T + v_{t-1},$$

то

$$\begin{aligned} M_{t-1,t} &= \mathcal{M}\{(f_{t-1} A^T + v_{t-1})^T (f_t A^T + v_t)\} = \\ &= A f_{t-1}^T f_t A^T + \mathcal{M}\{v_{t-1}^T\} f_t A^T + A f_{t-1}^T \mathcal{M}\{v_t\} + \mathcal{M}\{v_{t-1}^T v_t\} \end{aligned}$$

Учитывая предположение, что случайные отклонения для разных моментов времени независимы друг от друга, и имея в виду, что

$$\Phi_{t-1,t} = f_{t-1}^T f_t,$$

получаем зависимость между переходами показателей и переходами факторов

$$M_{t-1,t} = A \Phi_{t-1,t} A^T.$$

4) Чтобы найти обратную для матрицы переходов факторов, достаточно ее транспонировать.

Убедимся в этом. Умножим (3.4.1) слева на f_t^T , тогда получим :

$$f_t^T f_t = f_{t-1}^T f_{t-1} \Phi_{t-1,t}. \quad (3.4.3)$$

Учитывая ортогональность факторов, а также то, что произведение факторов для разных моментов времени задают матрицу переходов:

$$f_t^T f_{t-1} = \Phi_{t,t-1},$$

выражение (3.4.3) преобразуется к виду

$$I = \Phi_{t,t-1} \Phi_{t-1,t},$$

из которого следует

$$\Phi_{t,t-1} = \Phi_{t,t-1}^{-1}.$$

С другой стороны, транспонируем (3.4.1), после чего умножим справа на f_{t-1} , получим:

$$\begin{aligned} f_t^T f_{t-1} &= \Phi_{t-1,t}^T f_{t-1}^T f_{t-1}, \\ \Phi_{t,t-1} &= \Phi_{t-1,t}^T. \end{aligned}$$

После чего и получается, что

$$\Phi_{t-1,t}^T = \Phi_{t-1,t}^{-1}. \quad (3.4.4)$$

Таким же свойством обладают матрицы собственных векторов, соответствующих собственным значениям симметрической матрицы.

Покажем, что случайные отклонения уравнения факторной структуры для разных моментов времени являются независимыми:

$$\begin{aligned} \mathcal{M}\{v_{t-1}^T v_t\} &= \mathcal{M}\{(z_{t-1} - f_{t-1} A^T)^T (z_t - f_t A^T)\} = \\ &= \mathcal{M}\{z_{t-1}^T z_t - A f_{t-1}^T z_t - z_{t-1}^T f_t A^T + A f_{t-1}^T f_t A^T\}. \end{aligned} \quad (3.4.5)$$

Рассмотрим отдельные слагаемые последнего выражения. Математическое ожидание первого слагаемого имеет вид

$$\mathcal{M}\{z_{t-1}^T z_t\} = M_{t-1,t} = A \Phi_{t-1,t} A^T.$$

Пользуясь свойством (3.4.4), легко заметить, что из уравнения (3.4.1) следует

$$f_{t-1} = f_t \Phi_{t-1,t}^T,$$

или, если последнее выражение транспонировать, —

$$f_{t-1}^T = \Phi_{t-1,t} f_t^T,$$

и тогда математическое ожидание двух сомножителей второго слагаемого выражения (3.4.5) приобретает вид:

$$\mathcal{M}\{f_{t-1}^T z_t\} = \mathcal{M}\{\Phi_{t-1,t} f_t^T z_t\} = \Phi_{t-1,t} A^T.$$

Аналогично несложно заметить, что математическое ожидание двух сомножителей третьего слагаемого равно:

$$\mathcal{M}\{z_{t-1}^T f_t\} = A \Phi_{t-1,t}.$$

Относительно внутренних сомножителей четвертого слагаемого можно сказать, что их произведение равно матрице переходов факторов $\Phi_{t-1,t}$.

В итоге, выражение (3.4.5) приобретает вид

$$\mathcal{M}\{v_{t-1}^T v_t\} = A\Phi_{t-1,t}A^T - A\Phi_{t-1,t}A^T - A\Phi_{t-1,t}A^T + A\Phi_{t-1,t}A^T,$$

т.е. действительно случайные отклонения для разных моментов времени являются независимыми. Изначально это предполагалось, на этом предположении были выведены рассмотренные свойства, и данные свойства не противоречат предположениям.

Найдем оценку матрицы переходов факторов.

Найдем оценку $\Phi_{t-1,t}$ как решение экстремальной задачи поиска минимума суммы квадратов отклонений матрицы переходов значений переменных от матрицы переходов, задаваемых переходами значений факторов, т.е. в качестве основного выражения, формирующего критерий экстремальной задачи, является выражение:

$$M_{t-1,t} = A\Phi_{t-1,t}A^T,$$

и отличие левой и правой частей этого выражения, возведенное в квадрат, будет являться основным критерием. Следовательно, будем искать минимум функции

$$g_1(\Phi) = tr\{(M - A\Phi A^T)^T(M - A\Phi A^T)\},$$

где tr – «след матрицы», т.е. сумма диагональных элементов матрицы;

M и Φ – обозначают матрицы, соответственно, $M_{t-1,t}$ и $\Phi_{t-1,t}$, т.е. опущены нижние переходные индексы, поскольку другие переходы не рассматриваются, кроме как из предыдущего в последующее состояние.

Кроме того, оценки должны быть такими, чтобы удовлетворялось условие (3.4.4) – обратная матрица переходов может быть получена путем транспонирования.

Таким образом, имеет место задача поиска условного экстремума: найти минимум функции g_1 при ограничении (3.4.4). Для ее решения составим функцию Лагранжа

$$g(\Phi, \Lambda) = g_1(\Phi) + tr\{\Lambda(\Phi^T \Phi - I)\},$$

где Λ - матрица коэффициентов Лагранжа. Окончательно, функция Лагранжа имеет вид

$$g(\Phi, \Lambda) = \text{tr}\{(\mathbf{M} - \mathbf{A}\Phi\mathbf{A}^T)^T(\mathbf{M} - \mathbf{A}\Phi\mathbf{A}^T)\} + \text{tr}\{\Lambda(\Phi^T\Phi - \mathbf{I})\}. \quad (3.4.6)$$

Для того, чтобы найти условный экстремум, необходимо найти частные производные функции Лагранжа по ее аргументам Φ, Λ .

Частная производная функции g по матрице факторных переходов имеет вид:

$$\begin{aligned} \frac{\partial g}{\partial \Phi} &= \frac{\partial g}{\partial \Phi} (\text{tr}\{\mathbf{M}^T\mathbf{M} - \mathbf{M}^T\mathbf{A}\Phi\mathbf{A}^T - \mathbf{A}\Phi\mathbf{A}^T\mathbf{M} + \mathbf{A}\Phi^T\mathbf{A}^T\mathbf{A}\Phi\mathbf{A}^T\}) + \text{tr}\{\Lambda\Phi^T\Phi - \Lambda\} \\ &= \\ &= (-\mathbf{A}^T\mathbf{M}\otimes\mathbf{A}^T - \mathbf{E}(\mathbf{A}^T\otimes\mathbf{A}^T\mathbf{M}) + \mathbf{E}(\mathbf{A}^T\otimes\mathbf{A}^T\mathbf{A}\Phi\mathbf{A}^T) + \mathbf{A}^T\mathbf{A}\Phi\mathbf{A}^T\otimes\mathbf{A}^T)\text{vec}\{\mathbf{I}\} + \\ &+ (\mathbf{E}(\mathbf{A}^T\otimes\Phi) + \Phi\mathbf{A}^T\otimes\mathbf{I})\text{vec}\{\mathbf{I}\} = \text{vec}\{-\mathbf{A}^T\mathbf{M}\mathbf{A} - (\mathbf{A}\mathbf{M}^T\mathbf{A})^T + (\mathbf{A}^T\mathbf{A}\Phi^T\mathbf{A}^T\mathbf{A})^T + \\ &+ \mathbf{A}^T\mathbf{A}\Phi\mathbf{A}^T\mathbf{A}\} + \text{vec}\{\Phi\Lambda + \Phi\Lambda^T\} = \text{vec}\{-2\mathbf{A}^T\mathbf{M}\mathbf{A} + 2\mathbf{A}^T\mathbf{A}\Phi\mathbf{A}^T\mathbf{A} + 2\Phi\Lambda\}, \end{aligned} \quad (3.4.7)$$

где $\mathbf{E}$ – перестановочная матрица,

$\otimes$ – символ произведения Кронекера,

vec – операция «векторизации» матрицы, ее результатом является вектор-столбец, формируемый из исходной матрицы взятием строки за строкой.

В заключительном равенстве выражения (3.4.7) было учтено, что матрица коэффициентов Лагранжа – симметричная, т.е.

$$\Lambda^T = \Lambda,$$

поскольку уравнение

$$\Phi^T\Phi = \mathbf{I},$$

представляющее собой условие, является симметричной матрицей, и, следовательно, матрица коэффициентов Лагранжа должна быть также симметричной.

Найдем частные производные по Λ

$$\frac{\partial g}{\partial \Lambda} = (\mathbf{I}\otimes(\Phi^T\Phi - \mathbf{I})\text{vec}\{\mathbf{I}\} = \text{vec}\{\Phi^T\Phi - \mathbf{I}\}.$$

Приравнивая частные производные нулю, и переходя от векторной к обычной матричной форме, получим систему уравнений

$$\begin{cases} -A^T M A + A^T A \Phi A^T A + \Phi \Lambda = 0, \\ \Phi^T \Phi - I = 0. \end{cases} \quad (3.4.8)$$

Выразим Φ из первого уравнения системы (3.4.8), пользуясь свойствами векторного представления матриц:

$$\begin{aligned} (A^T A \otimes A^T A) \text{vec}\{\Phi\} + (I \otimes \Lambda) \text{vec}\{\Phi\} &= \text{vec}\{A^T M A\}, \\ (A^T A \otimes A^T A + I \otimes \Lambda) \text{vec}\{\Phi\} &= \text{vec}\{A^T M A\}, \\ \text{vec}\{\Phi\} &= (A^T A \otimes A^T A + I \otimes \Lambda)^{-1} \text{vec}\{A^T M A\}. \end{aligned}$$

Произведем небольшие преобразования, пользуясь, опять же, свойствами векторного представления матриц:

$$(\Phi \otimes I) \text{vec}\{I\} = (A^T A \otimes A^T A + I \otimes \Lambda)^{-1} (A^T M A \otimes I) \text{vec}\{I\}.$$

Перенесем правую часть последнего выражения налево и вынесем $\text{vec}\{I\}$ за скобки:

$$(\Phi \otimes I - (A^T A \otimes A^T A + I \otimes \Lambda)^{-1} (A^T M A \otimes I)) \text{vec}\{I\} = 0,$$

откуда

$$\Phi \otimes I = (A^T A \otimes A^T A + I \otimes \Lambda)^{-1} (A^T M A \otimes I).$$

Для последнего выражения найдем обратную матрицу

$$\begin{aligned} \Phi^{-1} \otimes I &= ((A^T M A)^{-1} \otimes I) (A^T A \otimes A^T A + I \otimes \Lambda), \\ \Phi^{-1} \otimes I &= (A^T M A)^{-1} A^T A \otimes A^T A + (A^T M A)^{-1} \otimes \Lambda, \end{aligned}$$

или после умножения обеих частей уравнения на $\text{vec}\{I\}$, и сворачивания в вектор, а затем, приравнивая аргументы оператора vec , получаем

$$\Phi^{-1} = (A^T M A)^{-1} A^T A A^T A + (A^T M A)^{-1} \Lambda.$$

Вынесем общий множитель правой части за скобки, получим

$$\Phi^{-1} = (A^T M A)^{-1} (A^T A A^T A + \Lambda). \quad (3.4.9)$$

Обращение выражения (3.4.9) дает матрицу факторных переходов, вычисленную с точностью до матрицы множителей Лагранжа:

$$\Phi = (A^T A A^T A + \Lambda)^{-1} A^T M A. \quad (3.4.10)$$

Если транспонировать эту матрицу, то получим

$$\Phi^T = A^T M^T A (A^T A A^T A + \Lambda)^{-1}. \quad (3.4.11)$$

Выражения (3.4.10) и (3.4.11) подставим в (3.4.8) и найдем матрицу Λ

$$A^T M^T A (A^T A A^T A + \Lambda)^{-2} A^T M A = I. \quad (3.4.12)$$

Из уравнения (3.4.12) выразим обратную матрицу:

$$(A^T A A^T A + \Lambda)^{-2} = (A^T M^T A)^{-1} (A^T M A)^{-1}. \quad (3.4.13)$$

Объединим обратные матрицы правой части выражения (3.4.13):

$$(A^T A A^T A + \Lambda)^{-2} = (A^T M A A^T M^T A)^{-1}, \quad (3.4.14)$$

и найдем обратную матрицу для выражения (3.4.14):

$$(A^T A A^T A + \Lambda)^2 = A^T M A A^T M^T A,$$

откуда получим значение матрицы коэффициентов Лагранжа:

$$\Lambda = (A^T M A A^T M^T A)^{1/2} - A^T A A^T A, \quad (3.4.15)$$

Найденное значение, т.е. выражение (3.4.15), подставим в уравнение (3.4.10) и окончательно получим

$$\Phi = (A^T M A A^T M^T A)^{1/2} A^T M A. \quad (3.4.16)$$

Следует отметить, что матрицы $A^T M A$ и $A^T M^T A$ не являются симметричными, поскольку матрица переходов M наблюдаемых переменных таковой не является. Однако произведение этих матриц есть матрица симметричная и для нее найти корень квадратный – несложно: достаточно извлечь квадратный корень из элементов диагональной матрицы собственных значений симметричной матрицы в ее разложении на произведение собственных векторов и собственных значений.

3.5. Факторный анализ в прогнозировании многомерных временных рядов

В отличие от обычной регрессионной зависимости, линейной или нелинейной, лаговое соотношение является наиболее приемлемым инструментом для формулирования прогнозного утверждения. Поскольку во многих ситуациях, когда рассматривается экономический процесс, описываемый несколькими переменными, связанными некоторым причинным

механизмом, есть основания полагать, что причины предшествуют следствиям. На практике одни события через некоторое время порождают другие, те в свою очередь становятся причинами следующих событий, и так создается цепь причинности. Так, повышение цен ведет к повышению социальных выплат, но повышение выплат в социальной сфере также может привести к повышению некоторых цен. Таким образом, повышение цен в некоторый период времени зависит от повышения цен в предшествующий момент времени, и это пример лаговой, авторегрессионной зависимости.

Когда исследуется система с сильными внутренними взаимосвязями, особенно характерными для экономики, одно уравнение для моделирования динамических процессов будет явно недостаточно. Чтобы построить достоверный прогноз необходимо объяснить два аспекта: первый – динамическое развитие системы, второй – взаимодействие различных переменных.

Поэтому, для решения вопросов, связанных с прогнозированием многомерных временных рядов используется модель, содержащая два уравнения, первое – уравнение переходов факторов, определяющих динамику системы:

$$\mathbf{f}_t = \mathbf{f}_{t-1}\mathbf{\Phi}_{t-1,t}, \quad (3.5.1)$$

и второе – уравнение определения значений экономических показателей на основе значений факторов, объясняющих взаимосвязи переменных:

$$\mathbf{z}_t = \mathbf{f}_t\mathbf{A}^T + \mathbf{v}_t. \quad (3.5.2)$$

Пусть на основании N наблюдений за переменными, представляющие собой матрицу $\mathbf{Z}$ – уже центрированную и нормированную матрицу реальных наблюдений, – найдены оценки факторных нагрузок $\mathbf{A}$, диагональной матрицы характеристик $\mathbf{D}^2$, матрицы переходов факторов $\mathbf{\Phi}$, и значений факторов для реальных наблюдений $\mathbf{F}$, которые определяются на основании уравнения

$$\mathbf{F} = \mathbf{ZA}(\mathbf{A}^T\mathbf{Z}^T\mathbf{ZA})^{-1/2}. \quad (3.5.3)$$

Тогда прогнозные оценки значений факторов для моментов в будущем $t = 1, 2, \dots, N_p$ (N_p – глубина прогноза) могут быть получены на основании уравнения (3.5.1), оценки значений показателей – на основании уравнения (3.5.2). Однако оценка значений факторов может быть уточнена на основании полученного прогнозного значения вектора переменных $\hat{\mathbf{z}}_t$ в момент времени t в будущем и оценки значения факторов, получаемой по уравнению (3.5.3) с добавленным в него, в качестве последней строки, вектором значений переменных $\hat{\mathbf{z}}_t$. Таким образом, для получения наиболее достоверной оценки значений факторов $\hat{\mathbf{f}}_t$ может быть использована техника фильтра Калмана.

Основные прогнозные уравнения имеют следующий вид

$$\hat{\mathbf{f}}_t = \hat{\mathbf{f}}_{t-1}\Phi + (\hat{\mathbf{z}}_{t-1} - \tilde{\mathbf{f}}_{t-1}\mathbf{A}^T)\mathbf{K}_{t-1} \quad (3.5.4)$$

и

$$\hat{\mathbf{z}}_t = \hat{\mathbf{f}}_t\mathbf{A}^T. \quad (3.5.5)$$

В последних двух уравнениях $\tilde{\mathbf{f}}_t$ – вектор уточняющих значений факторов, $\mathbf{K}_t$ – весовая матрица, $t - 1$ означает предшествующий момент времени. Алгоритм получения оценки прогнозных значений экономических показателей состоит из следующей последовательности вычислений. Прежде всего, задаются начальные значения матриц и векторов.

Начальные значения оценок переменных – это N -я строка наблюдений матрицы $\mathbf{Z}$:

$$\hat{\mathbf{z}}_0 = \mathbf{z}_N;$$

а вся матрица – $\mathbf{Z}_0 = \mathbf{Z}$; (напомним, что заглавной буквой обозначается матрица, т.е. таблица чисел, а прописной – вектор, или матрица-строка)

начальные значения уточняющих факторов – это N -я строка матрицы значений факторов $\mathbf{F}$:

$$\tilde{\mathbf{f}}_0 = \mathbf{f}_N;$$

начальное значение весовой матрицы

$$\mathbf{K}_0 = (\mathbf{A}\mathbf{A}^T + \mathbf{D}^2)^{-1}\mathbf{A};$$

начальное значение вспомогательной матрицы $\mathbf{P}$ – единичная матрица $\mathbf{I}$:

$$P_0 = I.$$

Далее, для моментов времени $t = 1, 2, \dots, N_p$ рекурсивно производятся следующие вычисления:

$$\left\{ \begin{array}{l} \hat{f}_t = \hat{f}_{t-1} \Phi + (\hat{z}_{t-1} - \tilde{f}_{t-1} A^T) K_{t-1}; \\ \hat{z}_t = \hat{f}_t A^T; \\ P'_t = \Phi P_{t-1} \Phi^T; \\ K_t = (A P'_t A^T + D^2)^{-1} A P'_t; \\ P_t = P'_t - P'_t A^T K_t; \\ Z_t = \begin{pmatrix} Z_{t-1} \\ \hat{z}_t \end{pmatrix}; \\ \tilde{f}_t = \hat{z}_t A (A^T Z_t^T Z_t A)^{-1/2}. \end{array} \right. \quad (3.5.6)$$

В блочной матрице Z_{N_p} , получаемой на последнем этапе вычислений, первые N строк – значения переменных, полученные в результате наблюдений, последние N_p строк – это прогнозные значения переменных. Эти значения – центрированные и нормированные. Чтобы получить реальные значения экономических переменных, необходимо преобразовать центрированно-нормированные значения по следующей формуле:

$$\hat{x}_{ti} = \bar{x}_i + \hat{z}_{ti} \sigma_i, \quad (3.5.7)$$

где $\hat{x}_{ti}$ – прогнозные значения i -й экономической переменной в реальной единицах измерения для t -го прогнозного момента времени;

$\bar{x}_i$ – среднее значение i -й экономической переменной в реальной единицах измерения;

$\hat{z}_{ti}$ – прогнозные значения i -й экономической переменной в безразмерных единицах измерения (центрированное и нормированное) для прогнозного момента времени t ;

σ_i – среднеквадратичное отклонение i -й переменной.

Рассмотрим практическое применение предлагаемой методики для решения задачи прогнозирования поведения экономических показателей. В подразделе 2.4 рассматривались проблемы получения факторного решения для модели исследования взаимосвязей изменения уровня процентных ставок с остатками на депозитных счетах по одному виду валюты. Теперь постараемся

исследовать связи остатков на депозитных счетах в национальной и иностранной валютах между собой, а также влияние на них уровня процентных ставок. Для этого рассмотрим соответствующую факторную модель, оценим ее параметры, а после чего построим прогноз на ближайшие пять месяцев – одна третья часть от имеющейся статистики. Теория прогнозирования рекомендует использовать в качестве глубины прогноза период, не превышающий десяти процентов от периода наблюдений за прогнозируемым процессом, в противном случае ошибка прогноза будет настолько значительной, что ему нельзя будет верить. Предлагаемая методика позволяет существенно увеличить глубину прогноза, однако строгое доказательство этого утверждения лежит за пределами рассматриваемых вопросов. Цель прогнозирования – проследить рекомендуемый моделью уровень процентных ставок, выяснить, как повлияет тенденция снижения процентных ставок по некоторым счетам на уровень остатков депозитов, и какое значение остатков на депозитных счетах следует ожидать на конец периода прогнозирования.

Рассмотрим факторную модель связи депозитов в различных валютах, в которой рассматривались следующие показатели:

X_1 – депозиты физических лиц в национальной валюте на первое число очередного месяца по счету 2630;

X_2 – депозиты физических лиц в национальной валюте на первое число очередного месяца по счету 2635;

X_3 – депозиты физических лиц в иностранной валюте на первое число очередного месяца по счету 2630;

X_4 – депозиты физических лиц в иностранной валюте на первое число очередного месяца по счету 2635;

X_5 – процентная ставка счета 2630 в национальной валюте;

X_6 – процентная ставка счета 2635 в национальной валюте;

X_7 – процентная ставка счета 2630 в иностранной валюте;

X_8 – процентная ставка счета 2635 в иностранной валюте.

Оценим матрицу факторных нагрузок и найдем значения факторов для факторной модели. Максимально правдоподобная оценка матрицы факторных нагрузок для решения при трех факторах получилась следующей (таблица 3.5.1):

Таблица 3.5.1 – Факторные нагрузки модели связи депозитов в национальной и иностранной валютах

	F_1	F_2	F_3
X_1	0,652	0,053	-0,062
X_2	0,287	0,324	-0,009
X_3	0,309	-0,140	0,883
X_4	0,319	0,260	-0,089
X_5	-0,158	-0,411	-0,065
X_6	-0,449	0,587	0,398
X_7	-0,249	-0,284	0,176
X_8	0,084	-0,459	0,125

Проведем интерпретацию факторного решения. В таблице 3.5.1 затенением выделены те переменные, которое наиболее нагружаемы данным фактором, т.е. в каждой строке выделялось наибольшее значение.

К первому фактору отошли две переменные, наиболее нагружаемые им, это – краткосрочные депозиты в национальной валюте и долгосрочные депозиты в иностранной валюте. Достаточно близкими значениями, имеющими также знак плюс, являются нагрузки при переменных X_2 и X_3 – долгосрочных депозитах в национальной валюте и краткосрочных депозитах в иностранной. Поскольку в абсолютной величине значений остатков на депозитных счетах заинтересовано именно данное кредитно-финансовое учреждение, то данный фактор можно интерпретировать как фактор активности банка по привлечению депозитов. Поскольку в исходных данных заметно, что уровень процентных ставок имеет тенденцию к снижению, кроме ставок по долгосрочным кредитам

в иностранной валюте, то первый фактор нагружает значения процентных ставок отрицательно.

Второй фактор оказался очень близок к генеральному фактору: к нему отошло наибольшее количество переменных, и, причем – все переменные, относящиеся к уровню процентных ставок. Таким образом, это фактор значимости процентных ставок. Но знак у переменных X_6 и X_7 – разный, а нагрузки при этих переменных наибольшие по сравнению с другими, следовательно, второй фактор может быть охарактеризован, как фактор предпочтения долгосрочных депозитов в национальной валюте.

Третий фактор оказался связанным, причем, с наибольшим значением нагрузки во всей матрице, с краткосрочными депозитами в иностранной валюте. Нагрузки при других переменных уровня депозитов – отрицательны и малозначимы. Т.е. для третьего фактора не интересна ни национальная валюта, ни длительность срока депозита, он отдает предпочтение иностранной валюте, размещенной на относительно небольшой срок. Если такую мотивацию связать с действиями не только банка, но и его клиентов, то такая тактика наиболее вероятна в кризисные моменты: отдать предпочтение более твердой валюте, но ненадолго, а отложить окончательное решение на потом. Таким образом, третий фактор – это фактор риска. Примечательно, что этот фактор интересуется процентными ставками на долгосрочных депозитах в национальной валюте (второе по величине значение нагрузки), но не до такой степени, чтобы эта переменная была отнесена как определяющая третьего фактора.

Значения факторов для пятнадцати месяцев у рассматриваемого банка оказались следующими (таблица 3.5.1):

Таблица 3.5.2 – Значения факторов модели связи депозитов в национальной и иностранной валютах

Месяц	F_1	F_2	F_3
1	-0,080	-0,353	-0,211

Продолжение таблицы 3.5.2

2	-0,026	-0,350	-0,198
3	0,031	-0,354	-0,157
4	0,253	-0,530	0,028
5	-0,292	-0,028	0,164
6	-0,422	0,069	0,033
7	-0,359	0,075	0,090
8	-0,032	0,029	0,491
9	0,083	0,046	0,603
10	-0,106	0,201	-0,025
11	-0,051	0,269	-0,262
12	-0,039	0,274	-0,32
13	0,047	0,271	-0,279
14	0,407	0,201	0,029
15	0,587	0,178	0,014

Отличительными особенностями значений факторов в течение рассматриваемого периода времени оказались следующие моменты. Фактор активности банка в значимой мере проявился в последние два месяца, до этого деятельность банка можно характеризовать, как неустойчивую: его положительные усилия чередовались с отрицательными. Наиболее отрицательные значения этот фактор имеет в 5-й, 6-й и 7-й месяцы, именно тогда произошел отток средств с депозитных счетов, что привело к уменьшению остатков в целом на всех депозитных счетах.

Второй фактор значимости процентных ставок характеризовался устойчивыми отрицательными значениями вплоть до шестого месяца. До него предпочтение вкладчики отдавали долгосрочным депозитам в иностранной валюте. На пятом месяце произошло резкое повышение процентных ставок на

долгосрочных депозитах в гривне, после чего значение второго фактора приобрело устойчивое положительное значение.

Безрисковые периоды отмечают значения третьего фактора. Это с 1-го по 3-й и с 10-го по 13-й месяцы. В остальные месяцы значение фактора – положительное, а ведь так имело место проявление фактора риска. Причем наибольшее его значение приходится на девятый месяц. Очевидно, такая динамика должна быть увязана с сопутствующими событиями. Но эта сторона анализа лежит за пределами данного исследования.

Среди трех факторов наиболее важным проявился второй фактор – значимости процентных ставок. На значимость указывает его близость к генеральному (тотальному) фактору. Его поведение может быть охарактеризовано как стабильно положительное.

Перед проведением процедуры прогноза на предварительном этапе получения оценок основных параметров факторной модели были найдены значения характеристик каждой переменной (напомним, что сумма значений характеристики и общности равна единице – значению диагонального элемента корреляционной матрицы). Значения характеристик D^2 определялись как диагональные элементы матрицы $R - AA^T$, где A – максимально правдоподобная оценка матрицы факторных нагрузок.

Приведем максимально правдоподобные оценки значений характеристик (таблица 3.5.3).

Таблица 3.5.3 – Значения характеристик переменных модели связи депозитов в национальной и иностранной валютах

Переменная	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Значение характеристики	0,569	0,813	0,106	0,823	0,802	0,295	0,826	0,766

На основании значений характеристик можно судить о вкладе самой переменной в общее значение ее дисперсии, а как следствие, информационную значимость самой переменной в формировании ее значений. Среди всех переменных с наименьшим значением характеристики, а соответственно,

наибольшим влиянием факторов следует выделить X_3 – краткосрочные депозиты в иностранной валюте (0,106 – значение характерности). Эта переменная в самой значительной степени связана с фактором риска, но и первый фактор также оказывает на нее заметное влияние. Тем не менее, изменение этой переменной может рассматриваться как индикатор риска: снижение остатков на краткосрочных депозитных счетах, или замедление их роста, будет свидетельствовать об уменьшении финансовых рисков, и, *contra versa*, неожиданный приток на эти счета будет говорить об усилении ожиданий кризисных явлений.

Следующая переменная с небольшим значением характерности переменная X_6 – процентные ставки по долгосрочным депозитам в национальной валюте, т.е. переменная достаточно сильно связана со всеми тремя факторами и с фактором активности банка по привлечению ресурсов, и с фактором привлекательности процентных ставок, и с фактором риска. Следовательно, процентные ставки по долгосрочным депозитам в гривне могут стать ключевыми в выборе оптимальной стратегии банка, как по привлечению депозитов, так и со снижением уровня риска.

Наименьшее влияние факторы оказывают на переменные X_7 – процентные ставки по краткосрочным депозитам в иностранной валюте (0,826 – значение характерности) и X_4 – долгосрочные депозиты в иностранной валюте (0,823 – значение характерности). Таким образом, значениями этих переменных банк может управлять в ограниченной мере.

Реализация алгоритма прогноза (3.5.6) на последующие 5 месяцев дала следующие результаты:

Таблица 3.5.4 – Прогноз значений переменных на последующие пять месяцев

Период	Переменные							
Месяц	Депозиты (тыс. грн.)				Процентные ставки (%)			
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
16	1970	7120	2238	4364	20,37	22,78	10,41	12,35
17	2050	8426	2216	4814	19,81	23,03	10,12	12,12
18	2153	10446	2183	5487	18,90	23,61	9,68	11,74
19	2274	13535	2139	6475	17,42	24,83	9,03	11,10
20	2398	18207	2096	7894	15,03	27,23	8,09	10,04

Как видно из прогнозных значений, процентные ставки продолжат тенденцию к снижению. Особенно заметное снижение будет происходить по ставкам краткосрочных депозитов в национальной валюте: за пять месяцев они снизятся на 5 процентов. Однако снижения остатков на этих счетах не произойдет. Также будут нарастать остатки на счетах долгосрочных депозитов в иностранной валюте при умеренном снижении процентных ставок. Лишь у краткосрочных депозитов в иностранной валюте произойдет уменьшение их уровня.

Прогнозная факторная модель рекомендует повышение процентных ставок по долгосрочным депозитам в национальной валюте, и при ставке на двадцатый месяц в размере 27 процентов годовых уровень остатков на этих счетах достигнет суммы в 18207 тыс. грн., т.е. именно эти депозиты обеспечат самый большой приток привлеченных средств.

Библиографический список

1. Математика и кибернетика в экономике: Словарь-справочник /М.: Экономика, 1971.–223 с.
2. Spearman C. General intelligence objectively determined and measured /C. Spearman.– Am. J. Psychoter, 1904. № 15. – p.201-293.
3. Cattell R. Handbook of multivariate experimental psychology /R. Cattell.– Chicago, 1966.–428 p.
4. Kendall M. Factor analysis. Part 1. Factor analysis as a statistical technique/ M. Kendall // J. Roy. Statist. Soc., 1950.– N 12.–p. 60-73.
5. Pearson K. On lines and planes of closest fit to systems of points in space / K. Pearson.–Philosophical Magazine., 1901.– N6.– p. 559-572.
6. Hotelling H. Analysis of a complex of statistical variables into principal components/ H. Hotelling.–Journal of Social Psychology, 1933.–N24.– p. 417-441.
7. Moulton F. The velocity of light.–Scientific Monthly, 1939.–N 48.–p.481-484.
8. Hotelling H. Analysis of a complex of statistical variables into principal components/ H. Hotelling.–Journal of Social Psychology, 1933.–N24.– p. 417-441.
9. Anderson T. Statistical inference in factor analysis /T. Anderson, H. Rubin.– Proc. Of the Third Symp. On Math. Statistics and Probability, 1956.–N5.– p. 111-150.
10. Thurstone L. Multiple factor analysis/ L. Thurstone.– Psychological Review, 1931.– N 38.– p. 406-427.
11. Tryon R. Cluster analysis/ R. Tryon.– Ann. Arbor: Edward Bros., 1939.– P. 122.
12. Хедли Дж. Линейная алгебра/ Дж. Хедли.– М.: Высш. шк., 1966.– 290 с.
13. Юл Дж. Теория статистики/Дж. Юл, М. Кендэлл.–М.: Госстатиздат, 1960.– 348 с.

14. Лоули Д. Факторный анализ как статистический метод/ Д. Лоули, А. Максвелл.— М.: Мир, 1964.— 224 с.
15. Хохлов В.В. Задачи и методы факторного анализа при моделировании сложных систем/ В.В. Хохлов // Рукопись.— Деп. В УкрНИИНТИ 25.02.93.— Севастополь, 1993.— № 273-Ук93.— 24 с.
16. Dwyer P. The contribution of an orthogonal multiple factor solution to multiple correlation / P. Dwyer.—Psychometrika, 1939.—N 4.— P. 163-171.
17. Guilford J. Psychometric methods/ J. Guilford. — N.Y., 1954.— P.216.
18. Spearman C. Theory of general factor/ C. Spearman.— British Journal of Psychology, 1946.— N 36.— P.117-131.
19. Иберла К. Факторный анализ/ К. Иберла.—М.: Статистика, 1980.—398с.
20. Kaiser H. A note on Tryon-Kaiser solution for the communalities/ H. Kaiser.—Psychometrika, 1989.—No 24.—P. 269-271.
21. Kaiser H. Solution for the communalities/H. Kaiser.— Research report, No.5,1996, Berkley, U. of California.— 11p.
22. Лапа В.Г. Математические основы кибернетики/ В.Г. Лапа.— К.: Вища школа, 1974.—458 с.
23. Харман Г. Современный факторный анализ/Г. Харман. — М.: Статистика, 1972.—486 с.
24. Бикел П. Математическая статистика/ П. Бикел, К. Доксам. —Вып. 2.—М.: Финансы и статистика, 1983.—254с.
25. Айвазян С.А. Прикладная статистика: Основы моделирования и первичная обработка данных/С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. — М.: Финансы и статистика, 1983.— 471 с., с.254.
26. Heermann E.F. Univocal or orthogonal estimators of orthogonal factors. Psychometrics, 1963, N 28. P.161-172.
27. Ansary A. Heterogeneous factor analysis model: a Bayesian approach /A. Ansary, K. Jedidi, L. Dube.— Psychometrika, 2002.—Vol. 67,No. 1.—P. 49-78.
28. М. Кендел. Временные ряды/Кендел М.— М.: Финансы и статистика, 1981.— 199 с.

29. Андерсон Т. Статистический анализ временных рядов /Т. Андерсон. – М.: Мир, 1976. – 756 с. С.14.
30. Бокс Дж. Анализ временных рядов: прогноз и управление /Дж. Бокс, Г. Дженкинс.–М.: Мир, 1974. –Вып.1.–408 с.
31. Brown R. Smoothing, forecasting and prediction of discrete time-series.–New Jersey: Prentice-Hall, 1982.–246 p.
32. Колмогоров А.Н. Интерполяция и экстраполяция стационарных случайных последовательностей. – Известия АН СССР, сер. Математика, 1941, т. 5, № 3.
33. Бриллинджер Д. Временные ряды: обработка данных и теория. – М.: Мир, 1980. – 536 с.
34. Хеннан Э. Многомерные временные ряды. – М.: Мир, 1974. – 575 с., с. 203.
35. Kalman R.E. New results in linear filtering and prediction theory. – Journal of Basic Engineering, 1961. – 83D.
36. Sorenson H.W. Kalman filtering techniques.–Advances in Control Systems Theory and Applications, 1966.– vol. 3.– pp. 219-292.

Монографія
Хохлов Володимир Володимирович

**БАГАТОМІРНИЙ ФАКТОРНИЙ АНАЛІЗ ЧАСОВИХ РЯДІВ БАНКІВСЬКИХ
ДЕПОЗИТІВ**

(російською мовою)

Монография
Хохлов Владимир Владимирович

**МНОГОМЕРНЫЙ ФАКТОРНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ БАНКОВСКИХ
ДЕПОЗИТОВ**

Коректор Л.П.Светлих
Нормоконтролер Г.М.Персідськов
Комп'ютерний набір і верстка В.В. Хохлов

Здано в набір 02.10.2008 р. Підп. до друку 28.04.09 ДК № 1272 від 17.03.03
Формат 60 x 90/16. Пап. тип. № 1. Офсет. друк. Ум. друк. арк. 12,75.
Тираж 100 прим. Зам. № __

Видавництво СевНТУ, м. Севастополь-53, вул. Університетська, 33, НМЦ,
т. (0692) 235-210
E-mail: root@sevgtu.sebastopol.ua; nms@sevgtu.sebastopol.ua

