

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ЦЕНТРАЛЬНЫЙ ЭКОНОМИКО-МАТЕМАТИЧЕСКИЙ ИНСТИТУТ

Г. Б. Клейнер С. А. Смоляк

ЭКОНОМЕТРИЧЕСКИЕ ЗАВИСИМОСТИ

**Принципы и методы
построения**



МОСКВА
«НАУКА»
2000

УДК 519.2
ББК 22.172
К 48

*Издание осуществлено при финансовой поддержке
Российского гуманитарного научного фонда (РГНФ)
проект № 99-02-16042*

Клейнер Г.Б., Смоляк С.А.

Эконометрические зависимости: принципы и методы построения. – М.: Наука, 2000. – 104 с.

ISBN 5-02-008348-8

Как установить зависимость между показателями экономических объектов, используя теоретическую информацию и данные об их наблюдаемых значениях? Можно сказать, что для этого существуют метод наименьших квадратов и другие методы, о которых говорится во всех учебниках по статистике. Между тем все не так просто.

Авторы книги – известные специалисты в области математического моделирования экономических объектов – предлагают использовать для построения экономико-статистических зависимостей *принцип максимальной согласованности*. Он позволяет полнее учесть разнородную исходную информацию о виде зависимости, результатах наблюдений и характере их ошибок и дает возможность решать многие задачи, недоступные обычным статистическим методам.

Для широкого круга экономистов, математиков, статистиков.

По сети АК

Научное издание

Клейнер Георгий Борисович
Смоляк Сергей Абрамович

**ЭКОНОМЕТРИЧЕСКИЕ ЗАВИСИМОСТИ:
принципы и методы построения**

Зав. редакцией *Р.С. Головина*. Редактор *В.С. Баковацкая*
Художник *Г.М. Коровина*. Художественный редактор *В.Ю. Яковлев*
Технический редактор *Е.Н. Ларкина*. Корректор *В.М. Ракитина*

Набор и верстка выполнены в издательстве на компьютерной технике

ЛР № 020297 от 23.06.1997

Подписано к печати 02.12.99

Формат 60 × 90 1/16. Гарнитура Таймс. Печать офсетная
Усл.печ.л. 6,5. Усл.кр.-отт. 6,9. Уч.-изд.л. 6,6. Тираж 700 экз. Тип. зак. 690

Издательство "Наука"
117864 ГСП-7, Москва Б-485, Профсоюзная ул., 90
Санкт-Петербургская типография "Наука"
199034, Санкт-Петербург В-34, 9-я линия, 12

ISBN 5-02-008348-8

© Издательство "Наука", 2000

Вычисления человека стали делаться все более акробатическими, математика перешла границы своего первоначального состояния и фактически превратилась в естественную часть мира, к которому она первоначально была лишь приложима. Вместо того, чтобы числа основывались на определенных явлениях, которым они соответствовали, ибо сами мы соответствовали понятной нам системе, весь мир постепенно стал основываться на числах, и, кажется, никого не удивляет тот факт, что сеть, некогда покрывавшая мир снаружи, превратилась в скелет, заключенный внутри.

В.В. Набоков

Одна из привлекательных особенностей экономической теории состоит в том, что с ее помощью всегда можно согласовать не только две, но и три различных точки зрения.

Р. Брейли, С. Майерс

ВВЕДЕНИЕ

Автор, разумеется, не уверен, что и эта книга не содер­жит какого-нибудь абсурда.

В. Тутубалин

Задача построения математической модели зависимости между экономическими показателями одного объекта или группы объектов рассматривается в рамках таких научных направлений, как регрессионный анализ [1, 2], детерминированный анализ показателей [3], теория производственных функций [4, 5] и др. Речь идет о построении ("отыскании", "оценивании", "восстановлении", "идентификации" и т.п. – термины меняются в зависимости от дисциплины, в которой ставится данная задача¹) функции $Y = f(X_1, \dots, X_n)$, отражающей зависимость "объясняемого" (результатирующего, зависимого) показателя Y некоторого экономического процесса P , протекающего на данном объекте, от "объясняющих" (независимых, экзогенных, предикторных) показателей X_1, \dots, X_n этого процесса, согласованной с исходной информацией и отвечающей целям исследования объекта. Желая подчеркнуть, что исходная информация включает в себя не только теоретические представления о протекании исследуемого процесса, но и "фактические данные" о наблюдениях за этим процессом, такие зависимости можно называть "статистическими", а применительно к экономическим процессам – эконометрическими. Для каждого из перечисленных направлений характерны свои подходы и сферы предполагаемого использования, ба-

¹ Термин "построение" нельзя считать однозначным, ибо под "построением зависимости" нередко понимается процедура, в которой вид зависимости между переменными выводится из некоторых более общих теоретических предпосылок. Типичным примером является построение квадратичной зависимости пройденного пути от времени в предположении постоянства ускорения.

зирующиеся на определенных предпосылках относительно процесса P , характера измерения или наблюдения показателей Y, X_1, \dots, X_n и связей между ними [6]. Различна также и трактовка самой построенной тем или иным способом зависимости $f(X_1, \dots, X_n)$.

Настоящая работа посвящена в основном теоретическим проблемам построения экономических зависимостей, достаточно хорошо согласованных с исходной информацией (с учетом ее характера и степени неопределенности). При этом неоднородность исходной информации обуславливает и многообразие постановок задач построения зависимостей. В данной работе мы пытаемся систематизировать известные и классифицировать возможные (нетрадиционные) варианты таких постановок, а также сопоставить результаты решения сходных задач в разных постановках. Это позволяет не только рассмотреть различные трактовки самого понятия о зависимости между характеристиками объектов, но и сформулировать некоторые общие принципы их практического оценивания. Особое внимание мы уделяем проблеме получения оптимальных (максимально согласованных с имеющейся информацией) оценок зависимости, что на практике сводится к оптимизации некоторых "критериев качества модели". Более широкие постановки задачи при этом в одних случаях приводят к новым критериям, а в других – к новым интерпретациям известных критериев оценивания параметров моделей. При этом иногда выясняется, что различные системы предпосылок, принимаемые при исследовании процессов, приводят к одинаковым или сходным критериям оценивания.

Четыре раздела книги посвящены соответственно определениям исходных понятий, принципам оценивания зависимостей, различным трактовкам понятия неопределенности и некоторым моделям оценки параметрических и непараметрических зависимостей, выходящих за рамки простых иллюстративных примеров.

Обобщая накопленный авторами опыт, книга тем не менее не предназначена к использованию в качестве учебного пособия. Для правильного понимания излагаемых вопросов требуется определенная математическая подготовка, включая теорию вероятностей (однако поверхностное знакомство с ее основами здесь может только помешать). Но главным требованием к нашим будущим читателям является их способность и готовность изменить свои устоявшиеся представления о тех зависимостях, с которыми они имеют дело, и о тех методах, которыми они пользуются при их построении. Представляется, что наличие таких качеств не зависит ни от сферы деятельности, ни от уровня подготовки специалиста, ни от наличия у него ученой степени. Поэтому, если попытаться определить, кому же предназначена эта книга, лучшим ответом будет: "тем, кто всерьез интересуется проблемой построения статистических зависимостей".

Ряд положений данной работы мы обсуждали с нашими коллегами по Центральному экономико-математическому институту РАН. Всем им, а в особенности академику В.Л. Макарову и доктору физико-математических наук С.А. Айвазяну – наша искренняя благодарность.

1. ОСНОВНЫЕ ПОНЯТИЯ

Если вам непонятно какое-то слово в техническом тексте, не обращайтесь на него внимания. Текст полностью сохраняет смысл и без него.

Закон Купера

1.1. Объекты и их характеристики

В практике экономических исследований проблема построения моделей зависимостей возникает обычно в следующем контексте [1, 7, 8 и др.]. Имеется некоторая группа экономических *объектов* или совокупность *состояний* одного и того же объекта – "генеральная совокупность" (иногда она именуется "предметной областью", а в теории измерений [9] для обозначения множества исследуемых объектов с учетом имеющейся на нем структуры отношений используется термин "эмпирическая область"). Относительно каждого объекта группы имеется определенная *информация*, характеризующая в различных аспектах его назначение, состояние, структуру или функции. В частности, имеется и информация, позволяющая считать некоторые из характеристик объекта зависимыми. Ставится задача: на основе всей имеющейся информации *оценить* (построить, восстановить) точно или приближенно эту зависимость. Процесс (и результат) такой оценки мы будем называть *моделированием* или *оценкой*.

Корректная постановка задач оценки зависимостей требует точного определения понятий "объект", "показатель", "наблюдение", "точность" и др. Однако это можно сделать, взяв за основу какие-то другие первичные понятия. Мы объявляем первичными (не определяемыми через другие) в первую очередь понятия "*генеральной совокупности*" и "*свойства*". Определения понятий "объект", "характеристика", "показатель" даются ниже, "наблюдение" и "точность" – в п. 1.2.

Генеральная совокупность (ГС) Ω рассматривается далее как множество, элементами которого являются *объекты* ω , в каком-то смысле однородные (например, предприятия, отрасли, рынки, жители города, технологические процессы). Под ω , как говорилось, может рассматриваться и определенное состояние реального объекта. Обычно имеют дело с конечной ГС, наблюдая ее отдельные объекты ω_i (обычно i – номер объекта или момента времени, в который объект наблюдается). "Правильное", адекватное целям исследования определение ГС (точнее, решение вопроса об отнесении конкретного объекта к ГС) является самостоятельной и не всегда тривиальной задачей. Так, было

бы ошибочным отнести завод к числу машиностроительных только потому, что это слово присутствует в его названии. Поэтому в ситуации, когда возможны грубые ошибки из-за неверного формирования ГС (см. п. 1.3), возникает задача выявления и учета таких ошибок, рассматриваемая в п. 4.3.

Объекты могут обладать различными *свойствами*. Примерами свойств являются "мощность", "возраст", "стабильность" и др. Обычно исследуются какие-то определенные свойства объектов, поэтому группировка объектов по одинаковым или близким свойствам во многом зависит от целей исследования. В этой части авторы частично разделяют, хотя и считают "экстремистской", точку зрения, высказанную еще в конце прошлого века, о том, что "объединение отдельных явлений в совокупности диктуется не особенностями материала, а целями изучения – теми точками зрения, с которыми статистик подступает к исследуемым явлениям" [10. С. 77–78]. Кроме целевых аспектов в основе каждого наблюдения лежит еще и некоторое априорное представление об объекте и возможных его свойствах (иначе просто было бы неясно, "за чем следует наблюдать" или к какому объекту относить полученную в ходе наблюдения информацию). Поэтому четкое определение ГС порой затруднительно.

Некоторым свойствам даны такие определения, в соответствии с которыми свойство может быть охарактеризовано (представлено) числом или другим *математическим объектом* (*математической конструкцией*) – парой чисел, булевой переменной, функцией и т.д. Подобные свойства будем называть **характеристиками**.

Характеристика объектов из Ω , выражаемая с помощью элементов некоторого множества V , тем самым, является отображением $\Omega \rightarrow V$ или функцией, заданной на множестве Ω и принимающей значения из множества V (далее мы несколько уточним это определение). Пространство всех подобных характеристик (измеримых отображений $\Omega \rightarrow V$) будем обозначать через $M(V)$. Если $V = \mathbf{R}$ – числовая прямая, то такие (числовые) характеристики объектов будем называть **показателями**.

Объекты, имеющие хотя бы одну несовпадающую характеристику, следует считать различными. Тем самым различными объектами будут не только завод и столовая, но и "завод в 1995 г." и "завод в 1996 г.". Однако во многих случаях завод и столовая могут быть отнесены к разным ГС, тогда как "завод в 1995 г." и "завод в 1996 г." – к одной ГС (хотя неясно, можно ли отнести к этой же ГС и "завод в 1896 г.").

Анализируя значения фиксированной характеристики для разных объектов ГС, исследователь часто делает вывод об их близости или, наоборот, отдаленности. В подобных случаях V является топологическим пространством. При этом появляется возможность говорить о сходимости некоторой последовательности характеристик к определенному пределу. Однако одного только наличия топологии в пространстве V , как правило, недостаточно. В ряде случаев исследователю нужно ко-

личественно оценить различия между характеристиками различных объектов (или между истинным и рассчитанным значением характеристики объекта). Для этого на множестве значений характеристик V часто вводится понятие *расстояния*. Тогда мы назовем множество V метризуемым и будем считать его подмножеством некоторого метрического пространства, а расстояние между точками X и Y из V обозначать через $\rho(X, Y)$.

Метрика на множестве Y естественно порождает "равномерную" метрику и в пространстве $M(V)$ соответствующих характеристик – при этом расстояние (мы будем обозначать его той же буквой ρ) между характеристиками $X(\bullet)$ и $Y(\bullet)$ определяется как максимально возможное для объектов ГС расстояние² между значениями этих характеристик по всем объектам Ω : $\rho((X, Y) = \sup_{\omega \in \Omega} \rho(X(\omega), Y(\omega))$. Близкими в этом смысле

(по крайней мере для совокупности стабильно работающих предприятий) являются показатели среднегодовой и среднесписочной численности работников предприятия.

1.2. Наблюдения, измерения, оценки

В основе формирования информационной базы построения статистических зависимостей лежат такие ключевые понятия прикладной статистики, как *наблюдение, измерение и оценка*. В различных работах даются разные определения этих понятий. Для уточнения и формулировки используемых в данной работе определений примем некоторое соглашение: условимся, что под словами "реально выполнимая процедура" будем понимать некоторые действия, выполнение которых априорно не кажется никому невозможным. Если речь идет об исследовании формы поверхности Юпитера, то будем считать, что это реально выполнимая процедура.

Под **наблюдением** теперь будем понимать реально выполнимую процедуру сбора и фиксации информации об объекте. **Измерение** (характеристики) – это также реально выполнимая процедура, позволяющая *однозначно и точно* определить значение этой характеристики в заданных условиях. Характеристика, допускающая измерение, называется **измеримой**. **Оценкой** характеристики³ объекта называется некоторое значение (другой или той же самой) измеримой характеристики этого объекта. При этом оценка может принадлежать тому же

² Иные метрики в пространстве $M(V)$, ориентированные не на максимальное, а на "среднее" расстояние, по-видимому, в общем случае будут неестественны, поскольку на Ω (а это множество обычно либо конечно, либо счетно) априорно не предполагается существование какой-либо меры, необходимой для исчисления "среднего расстояния" между его объектами (да и ввести такую меру довольно затруднительно).

³ Правильнее было бы говорить об оценке не самой характеристики, а ее *значения*. Однако там, где это не искажает содержания высказывания, мы будем говорить просто "характеристика".

множеству V , что и оцениваемая характеристика (далее мы в основном рассматриваем именно такие, *точечные* оценки), но может и не принадлежать ему.

Таким образом, в отличие от наблюдения, результатом которого может быть *любая* информация о характеристике объекта, результатом измерения должна быть *точная* информация об этой характеристике. При этом сама характеристика рассматривается как объективно присущая объекту, независимо от того, можно ли как-нибудь ее измерить. Это позволяет трактовать значения характеристик как "истинные", разумеется, имея в виду, что истинное для одних условий значение может стать ошибочным для других.

По отношению к наблюдению можно говорить о целях и результатах. **Целью** наблюдения в общем случае является получение информации относительно некоторого набора X_1, \dots, X_n характеристик объекта. Мы будем считать, что этот набор конечен, и в отдельных случаях рассматривать его как одну (векторную, обозначаемую жирным шрифтом) характеристику X .

Результатом наблюдения H также является некоторая (измеримая) характеристика объекта. В статистике обычно считается, что результатом наблюдения показателя является точное или приближенное значение *этого же показателя*, а потому само наблюдение есть процедура его (точного или неточного) измерения. Такая трактовка представляется слишком узкой. Во-первых, процедуры измерения целесообразно рассматривать как точные, используя для неточных измерений термин "*оценка*". Во-вторых, результаты наблюдения могут и не являться измерениями каких-либо характеристик. Наконец, если наблюдение проводится с целью оценки значения какой-либо характеристики из $M(V)$, то результатом наблюдения может и не быть элемент множества V (см. ниже примеры 1.5–1.11). С этих позиций наблюдение удобно трактовать более широко – как процедуру получения и фиксации *какой-либо* информации об объекте. При этом получаемая информация – результат наблюдения – является измеримой функцией на множестве объектов $\Omega = \{\omega\}$. В частности, наблюдением характеристики является ее *оценка*. Если результатом наблюдения является истинное значение характеристики, такое наблюдение назовем *точным*.

Измеримость характеристики мы связываем с реальной выполнимостью процедуры измерения. При этом понятие "реальной выполнимости" мы считаем первичным, определяемым в каждой конкретной задаче на содержательном уровне (это не исключает, разумеется, построения каких-либо теорий "практической измеримости"). Вопрос о том, какие процедуры считать измерениями, а какие – нет (какие значения характеристики, полученные путем применения этой процедуры, считать точными, а какие – неточными), нетривиален. Ответ на него зависит и от структуры множества V , и от формы представления результатов измерения, и от знаний и опыта исследователя. Приведем несколько примеров.

Пример 1.1. Основные фонды любого предприятия подразделяются на активные и пассивные. "Активность" конкретного средства труда, относящегося к основным фондам, определяется тем самым принадлежностью его к одному из двух указанных видов. Измерением будет только такая процедура, которая "правильно" отнесет наблюдаемый объект основных фондов к тому или иному виду. ■

Пример 1.2. Для определения стоимости строительства объекта строительными предприятиями составляется смета. Сильно огрубляя порядок сметного ценообразования, можно сказать, что в смете определяется состав и объем всех необходимых работ и рассчитывается (с помощью специальных нормативов) сметная стоимость каждой из них в отдельности, а затем – всех вместе. Точно определение сметной стоимости означает, что в смете отражены все необходимые работы, причем объем и стоимость каждой из них указаны точно. Однако обычно в процессе строительства с учетом выясняемых на месте конкретных условий строительства объемы указанных в смете работ уточняются, одновременно выявляется необходимость выполнения дополнительных ("пропущенных в смете") работ. Поэтому точно определить сметную стоимость строительства на этапе разработки проекта невозможно (равно как и вообще любое измерение показателей, относящихся к неопределенному будущему), однако это становится невозможным после завершения строительства на базе так называемой "исполнительной сметы", куда внесены все дополнения и уточнения. ■

Таким образом, понятие реальной выполнимости процедуры измерения является определенной идеализацией, оправданной и практически полезной в ситуациях, когда ошибки измерения достаточно малы. С другой стороны, есть и "принципиально неизмеримые" (хотя и допускающие оценку) характеристики. К ним в первую очередь относятся вероятности тех или иных событий (например, стихийных бедствий или отказов оборудования). Неизмеримы, по нашему мнению, функции полезности экономических субъектов, платежеспособный спрос на продукцию, предельные затраты и ряд других характеристик экономических процессов и явлений, явно фигурирующих в разных экономико-математических моделях. Неизмеримыми в определенном смысле можно считать и такие физические характеристики объектов, точное измерение которых осуществляется методами разрушающего контроля (скажем, точно измерить предельную нагрузку, которую может выдержать данный мост, можно только, разрушив его), поскольку после проведенного измерения объект измерения перестает существовать. Неизмеримые показатели уместно называть **параметрами**. Это не противоречит обычной трактовке параметров как коэффициентов в подлежащих определению зависимостях: подобные задачи представляют интерес только тогда, когда входящие в них параметры неизмеримы, т.е. не могут быть однозначно и точно определены.

Неизмеримость характеристики, конечно, затрудняет получение информации о ней, но не мешает ее оценивать, например, измеряя вместо

нее заменяющие, в каком-то смысле близкие характеристики. Такая процедура, при которой непосредственно не измеримая характеристика заменяется другой (индикатором), значения которой можно непосредственно измерить, называется индикацией [11]. Иными словами, при этом способе оценки вместо измерения того, что *нужно* измерять, измеряется то, что реально *можно* измерить (это аналогично поиску утерянной вещи не в том месте, где она потеряна, а под ближайшим фонарем). У одной характеристики может быть несколько индикаторов. Так, индикаторами изменения физических объемов производства промышленной продукции может быть не только объем продаж, скорректированный на индекс цен, но и объем потребления электроэнергии – это два "принципиально разных" индикатора, дающих обычно разные оценки. Другой способ оценки состоит в использовании иных методов и процедур измерения, т.е. в переходе от реально невыполнимых к реально выполнимым процедурам. В этом случае результат оценки может отличаться от истинного значения характеристики за счет "несовершенства измерительных приборов (или методов)". Необходимость в оценивании может возникнуть и при использовании, казалось бы, точных данных статистической отчетности при анализе рядов динамики, например, если в анализируемом периоде изменяются правила ("технология") составления статистической отчетности. При этом "стыковка" данных разных лет осуществляется путем введения соответствующих корректирующих коэффициентов, оценивающих влияние фактора смены методологии. Другим примером изменения "технологии" измерений может быть рост числа зарегистрированных нарушений трудовой дисциплины на предприятии с приходом нового директора и соответственным ужесточением требований к дисциплине [11]. Приведем примеры оценок ряда показателей.

Пример 1.3. Многие дискретные и непрерывные модели содержат такую характеристику предприятия, как средняя за период численность персонала. По своему смыслу за период T (например, за год) она должна определяться как среднее из численности работников по дням отчетного периода. В отчетности такого показателя нет, но хорошей его оценкой может быть индикатор *среднесписочной* численности (для месяца – полусумма численностей на начало и конец месяца, для года – среднее арифметическое из среднесписочных численностей по месяцам). ■

Пример 1.4. По большой группе предприятий за 1980–1995 гг. наблюдается ряд показателей, в том числе стоимость основных средств на конец года, содержащаяся в годовом балансе. Выяснилось, что по одному из предприятий данные об этом показателе за 1989 г. отсутствуют. Однако, как видно из показателей других предприятий группы, стоимость основных средств на конец 1989 г. близка к полусумме значений этой стоимости за 1988 и 1990 г. Применяв этот способ к данному предприятию, получим разумную оценку "пропущенной" характеристики. ■

Проведенное в наших определениях разграничение понятий "наблюдения" и "измерения" обусловлено разнообразием форм информа-

ции, получаемой в результате наблюдений. Как уже говорилось, при наблюдении характеристики из $M(V)$ результат наблюдения может выражаться элементом не множества V , а некоторого более широкого "пространства наблюдений" W (иногда в таких случаях говорят об оценке одних характеристик по результатам измерения других; задачами такого рода занимается геометрия – так, задача 46 в [12] сводится к оценке площади выпуклой фигуры по длинам ее проекций на четыре прямые). Приведем ряд примеров.

Пример 1.5. Четыре эксперта дают обоснованные экспертные оценки значению некоторого показателя. Здесь наблюдением следует считать всю процедуру сбора и фиксации экспертной информации, однако это – *неоднозначное наблюдение*, ибо его результатом являются четыре числа, а не одно. Аналогичная ситуация возникает, когда для одного и того же показателя в разных формах отчетности предприятия приводятся разные (хотя и близкие) значения, либо когда результаты наблюдений даются не числами, а интервалами их изменения. Элементами пространства W здесь удобно считать произвольные (конечные или бесконечные) подмножества числовой оси. ■

Пример 1.6. Пусть по данным измерительного прибора высота объекта (измеримая характеристика, показатель) определена в 25 см с точностью до 1 см. Результатом такого наблюдения является уже не число, а пара чисел (25 и 1). ■

Пример 1.7. Количество L жителей города, предпочитающих эски-мо другим видам мороженого, оценивалось по данным выборочных опросов. С учетом возможных ошибок такого метода социологи говорят, что число L лежит в таких-то пределах с вероятностью 90%, в таких-то пределах – с вероятностью 80% и т.д. (в математической статистике в этой ситуации вместо вероятности говорят о степени или уровне доверия, однако различие в данном случае несущественно). При этом результатом наблюдения оказывается не число, а распределение вероятностей (степени доверия). ■

1.3. Ошибки наблюдений

Ясно, что наблюдения характеристики производятся с целью ее оценки, и обратно – любая оценка является результатом какого-то наблюдения. Однако произвольная оценка может нести мало информации о нужной характеристике или вообще не нести никакой информации о ней. В этой связи можно говорить об **ошибках наблюдения**, понимая под ними несоответствия между наблюдаемыми значениями характеристик и их истинными значениями. Под данную трактовку подпадают и ошибки *регистрации*, возникающие при искажении (умышленном или неумышленном) результата наблюдения при его фиксации. Примерами здесь могут служить неверное считывание показаний регистрирующего прибора, запись числа не в ту клетку анали-

тической таблицы или неправильная установка десятичной запятой при записи этого числа.

Если некоторая процедура наблюдения позволяет достаточно точно оценить характеристику любого объекта совокупности, такую процедуру естественно рассматривать как достаточно точную. Отметим, что точность процедуры зависит:

1) от состава ГС. Если рассматривать одни и те же наблюдаемые объекты как элементы разных ГС, то один и тот же метод наблюдения будет охарактеризован разной точностью. Так, метод установления среднегодовой стоимости основных средств как полусуммы значений этой стоимости на начало и конец года будет достаточно точным в годы, когда темпы инфляции невысоки, но может давать большие ошибки в периоды высокой инфляции;

2) от состава наблюдаемых объектов (если наблюдается лишь часть объектов ГС). Поэтому, например, нельзя дать однозначную оценку точности методов выборочных социологических обследований населения – в лучшем случае можно лишь сказать, в скольких процентах случаев эти обследования давали такую-то ошибку.

Некоторые характеристики объектов, например, длину или массу, нельзя измерить абсолютно точно, хотя имеется возможность повышать точность измерений. Для идеализированного описания подобных характеристик удобно ввести следующее определение. Характеристика X называется *асимптотически измеримой*, если существует сходящаяся к ней в топологии пространства $M(W)$ последовательность оценок $H^k \rightarrow X$ при $k \rightarrow \infty$. Включение в рассмотрение асимптотически измеримых характеристик расширяет круг как инструментов, так и объектов экономических исследований.

Основная причина ошибок наблюдения – невозможность (теоретическая или практическая) измерения характеристики, т.е. применения "приложенной к ее определению" методики измерения. При таком подходе ошибки рассматриваются как "неизбежное зло", а во главу угла ставится не "борьба" с ними и не стремление их элиминировать, а максимальное использование полученной в результате наблюдений информации для реализации поставленных при измерении целей. Некоторые технические и психологические аспекты проблемы возникновения ошибок и борьбы с ними, оставшиеся за пределами нашего исследования, проясняются в оригинальной работе [13].

В зависимости от формы проявления можно выделить два типа ошибок. Ошибки *первого* типа возникают, когда и сама характеристика, и результат ее наблюдения принадлежат одному и тому же множеству V (см. например 1.3). В том случае, когда результат реального наблюдения принадлежит не множеству V , а более широкому пространству W , можно говорить об ошибках *второго* типа. Примером является ситуация, когда результатом наблюдения показателя оказывается не число, а интервал на числовой прямой или вероятностное распределение (см. пример 1.7).

Ввиду их важности и распространенности следует указать три не

подпадающих под приведенное определение вида ошибок, часто встречающиеся в практике эконометрических исследований: ошибки классификации, ошибки спецификации и ошибки модели.

Ошибки классификации возникают, когда в состав ГС не включены объекты, которые на самом деле (т.е. в соответствии с имеющейся информацией) к ней относятся. Например, исследуется совокупность промышленных предприятий, однако в ГС и в число объектов наблюдения не включены предприятия промышленности строительных конструкций и деталей, которые обычно учитываются не по отрасли "Промышленность", а по отрасли "Строительство". Та же проблема возникает при формировании чистых отраслей в межотраслевом балансе.

Ошибки спецификации имеют место при неправильном отборе объектов (включении в ГС таких объектов, которые на самом деле к ней не относятся) или факторов, влияющих на исследуемую характеристику (объясняющих переменных). Такие ошибки (например, использование сводных данных по крупной строительной организации, включающей кирпичный завод и автобазу) часто обусловлены недостатками в теоретическом анализе рассматриваемых явлений.

Ошибки модели проявляются в неправильном выборе вида искомой зависимости (например, отыскивается линейная зависимость между характеристиками в то время, как на самом деле они связаны степенной зависимостью). По существу, здесь также идет речь об ошибках спецификации, однако специфицируется здесь не состав аргументов, а вид модели. Часто эти два вида ошибок тесно связаны между собой. В рамках нашего подхода указанные ошибки не являются ошибками наблюдения, а возникают на стадии, предшествующей обработке результатов наблюдений (при выборе ГС, совокупности исследуемых характеристик, теоретическом анализе зависимостей между ними).

Наблюдения, как правило, характеризуются *неопределенностью*, под которой мы понимаем *неполноту* или *неточность информации о значениях наблюдаемой характеристики*. При этом неполнота определяется по отношению к тому (порой неизвестному) объему информации, который позволяет полностью определить истинное значение характеристики, а неточность трактуется как расхождение между истинными и полученными в ходе наблюдения данными. Неопределенность присуща практически любым наблюдениям, но мера, степень и конкретные формы ее проявления могут быть различными. В то же время там, где неопределенность незначительна, оказывается полезной такая абстракция, как "отсутствие неопределенности" – *детерминированность*. Так, если измерение длины предмета производится каким-то измерительным прибором, мы не можем заранее сказать, что результатом измерения будет истинная длина этого предмета. Здесь можно утверждать лишь, что ошибка в измерении длины будет лежать в таких-то пределах. Результат измерения будет

детерминирован, если измерение производится точным прибором (скажем, детерминированным может быть измерение количества крупных предметов, находящихся в некотором помещении). При использовании экспертных оценок для измерения показателя неопределенность проявится в несовпадении этих оценок. Здесь трудно говорить о детерминированном результате измерения, хотя нечто близкое к нему возможно, когда многие независимые аспекты дают одну и ту же оценку чему-либо.

Мы определили неопределенность как неполноту и неточность информации. С этих позиций информацию о наблюдаемых характеристиках объектов можно разделить на четыре вида: *полная и точная*; *неполная и точная*; *полная и неточная*; *неполная и неточная*. Различие между этими видами информации проиллюстрируем на примере наблюдения объема реализованной профильной продукции хлебозавода:

- *полная и точная*. Так будет при точном измерении физических объемов реализованного количества буханок хлеба различных сортов;

- *неполная и точная*. Этот случай имеет место при точном измерении только общего объема реализованного хлеба (без разбивки по сортам);

- *полная и неточная*. Этот случай имеет место, когда объемы реализации разных сортов хлеба измерялись, но измерения произведены неточно;

- *неполная и неточная*. Этот случай имеет место при неточном измерении общего объема реализации хлебопродукции (без разбивки по сортам).

Заметим теперь, что истинное значение характеристики, в отличие от результата его наблюдения, исследователю обычно неизвестно и потому трактуется как неопределенная (и часто неизмеримая в смысле п. 1.2) величина. Между тем моделирование взаимосвязи между характеристиками объекта и результатами наблюдений осуществляется с иных позиций: первые считаются определенными (хотя и неизвестными) величинами, тогда как вторые – неопределенными, значения которых могут быть различными в зависимости от условий и методики наблюдений. Другими словами, результат наблюдения трактуется как реализация неопределенной величины, одно из возможных ее значений. Сущность моделирования состоит при этом в формализации и учете неопределенного характера результатов наблюдений.

Таким образом, первичным является *неопределенный характер* результатов наблюдения, а вторичным – представление этих результатов в виде некоторой функции (например, суммы или произведения) от *истинного* значения измеряемой характеристики и *ошибки* наблюдения. С этих позиций целесообразно говорить об ошибке наблюдения только тогда, когда из характера неопределенности вытекает возможность соответствующего представления. Так, если

результат наблюдения – случайная величина, имеющая нормальное распределение, и ее средним является истинное значение наблюдаемой величины, вполне корректно говорить о наличии аддитивной нормально распределенной ошибки с нулевым средним. С другой стороны, если результат наблюдения имеет экспоненциальное распределение и его среднее значение совпадает с истинным значением наблюдаемой величины, разделение на "истинную часть" и ошибку становится искусственным и не упрощает анализа. Отметим, что в справочнике [14] регрессия величины Y по отношению к факторам X_1, \dots, X_n определяется как функция $Y = g(X_1, \dots, X_n) + h(Y, X_1, \dots, X_n)$, где $h(Y, X_1, \dots, X_n)$ рассматривается как поправочный член, так что и здесь разделение моделируемой величины на "истинную часть" и "ошибку" выглядит произвольным. Поэтому мы будем использовать понятие "ошибка наблюдения" только там, где результат наблюдения допускает естественное, т.е. объективно обусловленное ситуацией, разложение на "истинную часть" и ошибку.

Обратим внимание также на то, что само по себе признание какого-либо результата наблюдения неопределенным не несет никакой информации и может быть приложено к любой ситуации (поскольку даже точно измеренные, детерминированные показатели могут быть рассмотрены как частный случай неопределенных). Упоминание о неопределенности становится содержательным и требующим учета лишь тогда, когда конкретизирован ее вид (характер). Иными словами, *неопределенность должна рассматриваться не как синоним отсутствия информации, а как указание на специфический вид имеющейся информации, не сводящийся к указанию конкретного численного значения той или иной характеристики*. Сама по себе констатация факта наличия неопределенности, неточности или неполноты информации об условиях и результатах наблюдений не является основанием для принятия каких-либо модельных решений, подобные основания появляются лишь после и на основе конкретизации вида неопределенности и характера возможных ее проявлений. С этих позиций представляются некорректными, например, постановки задачи наилучшей оценки параметра в параметрической зависимости в условиях, когда результаты наблюдений являются неопределенными, но о "виде" этой неопределенности ничего не известно.

1.4. Зависимости между характеристиками объектов

Все объекты ГС предполагаются *целостными*, поэтому естественно, что любые их характеристики в общем случае должны быть взаимозависимыми – изменение одних должно как-то сказываться на изменении других. Однако в каждый данный момент времени характеристики данного объекта фиксированы, и говорить об их изменении и о влиянии этих изменений на другие характеристики объекта можно только, рассматривая объект как представи-

теля соответствующей ГС, только в рамках которой и имеет смысл соответствующая зависимость⁴.

Зависимости между характеристиками Y и X объекта моделируются обычно в виде функции $Y = f(X)$. "Стандартная" трактовка этого равенства означает, что каждому возможному значению X отвечает единственное определенное значение Y . При такой трактовке сама построенная зависимость, т.е. функция f , также может рассматриваться как *характеристика*, но уже не объекта, а всей ГС. Следовательно, к зависимостям применимы все понятия и определения, введенные и обсуждавшиеся в п. 1.2. В частности, можно говорить о генеральной совокупности зависимостей $\{f\}$, понимая под этим класс функций, объединенных какими-либо общими свойствами (например, свойством быть моделями одного и того же объекта или свойством линейности). Если функция f – полином, то его коэффициенты в свою очередь должны трактоваться как *параметры*, характеризующие ГС (а если предложена процедура их измерения, то и как *показатели* этой ГС). С этих позиций мы имеем дело с задачей возможно более точной оценки таких характеристик ГС, как функция f или ее параметры, по (обычно неполным и неточным) данным наблюдений отдельных объектов ГС. Соответственно мы можем говорить об "*истинной*" зависимости (в процессе оценки она рассматривается еще и как *искомая*) между показателями объекта и о *рассчитанной* зависимости, найденной в ходе решения задачи (оценку, полученную при использовании наилучших в каком-либо смысле методов, будем называть также "оптимальной").

Мы будем считать, что теоретический анализ, позволяющий выявить сам факт наличия зависимости, а в отдельных случаях – установить ее характер (вид, свойства), уже выполнен. Примеры такого анализа дает теория производственных функций [4, 6] или модели развития производства [15, 16]. Тем самым не ставится под сомнение ни отбор объясняемой и объясняющих переменных, ни существование искомой зависимости, ни характер (вид) этой зависимости, если он должным образом теоретически обоснован. Однако обычно установление зависимостей затрудняется:

- большим числом объясняющих переменных ("проклятие размерности");
- ошибками наблюдения переменных;
- неполной информацией о виде зависимости⁵ (см. пример 1.8);

⁴ Помимо рассматриваемых в книге зависимостей между характеристиками *одного* объекта есть и такие, которые связывают характеристики *одного объекта* в разные моменты времени или *разных объектов*. Примером является зависимость стоимости акций данной фирмы от положения на финансовом рынке страны, измеряемого ожидаемой отдачи всего рыночного портфеля акций. Методы оценки таких зависимостей аналогичны.

⁵ Обычно содержательный анализ экономических процессов позволяет установить, что соответствующим зависимостям между показателями присущи какие-либо из следующих свойств: линейность (прямая пропорциональность), монотонность, унимодальность, выпуклость или вогнутость и т.п. Нередко теория указывает на сепар-

- ограниченностью количества наблюдений.

Поэтому обычно результаты оценки зависимостей носят приближенный характер.

Имеются различные постановки задач оценки зависимостей, и некоторым из них, казалось бы, отвечают иные, нестандартные трактовки понятия зависимости. Ниже мы рассмотрим ряд таких постановок и убедимся, что это не так.

1.5. Различные типы статистических зависимостей

1.5.1. "Точные" зависимости. Задача интерполяции

Обычно, говоря о неточности оценки статистической зависимости, имеют в виду влияние ошибок наблюдений. Во многих случаях это действительно так, однако некоторые проблемы принципиального характера выявляются уже на примере задачи интерполяции – оценки неизвестной функции $Y = f(X)$ по T заданным (без ошибок) точкам ее графика $(X_1, Y_1), \dots, (X_T, Y_T)$. Эта постановка задачи может рассматриваться как в определенном смысле "минимальная" – исходная информация об объекте представлена здесь только значениями функции в наблюдавшихся точках, наблюдения которых предполагаются точными. Искомая зависимость $Y = f(X)$ при этом "точная", отражает однозначное соответствие между Y и X и потому трактуется стандартно. Отметим, что хотя задачи интерполяции часто встречаются в инженерной практике, в прикладной статистике им внимания не уделяется. Обычно это объясняется тем, что выявить функциональную зависимость между переменными можно, используя методы "детерминированного анализа". Между тем никаких общих, универсальных методов выявления детерминированных зависимостей не существует. Приведем два примера.

Пример 1.8. Несмотря на огромное количество точных данных, до сих пор неизвестно, какова зависимость простого числа от его порядкового номера. ■

Пример 1.9. Эффективность инвестиционного проекта часто рассчитывается по "закрытым" компьютерным программам ("КОМФАР" и др.). Зависимости показателей эффективности от нормы дисконта или темпа инфляции при этом устанавливаются вариантными расчетами – аналитические их выражения здесь неизвестны. ■

Особенность задачи интерполяции в том, что ее решением может

рабельность зависимости, что сводит задачу к оценке нескольких функций от меньшего числа или даже от одной переменной. Наконец, довольно часто влияние всех или некоторых объясняющих переменных описывается линейными зависимостями, хотя это не всегда должным образом обосновывается, а иногда и противоречит результатам других исследований. Вообще, вопросам теоретического обоснования вида и характера зависимостей между экономическими характеристиками посвящено не так уж много работ, и эти вопросы заслуживают самостоятельного и более детального рассмотрения.

быть любая функция, график которых проходит через все точки (X_i, Y_i) , т.е. такая, для которой отклонения заданных (трактуемых как наблюдаемые) значений от "рассчитанных по результатам оценки" равны нулю. Такое решение задачи, однако, малопригодно в практических целях. Более содержательными и имеющими "разумные" решения оказываются постановки, в которых "минимальная" исходная информация дополняется требованиями к функции, вытекающими из предположений о ее характере (см. пп. 2.4 и 2.8). Мы видим, что даже в ситуации, когда известно о "точном" характере зависимости между переменными, "точно" восстановить ее не удастся. Иными словами, "точный" характер *искомой зависимости* не противоречит приближенному характеру ее *оценки*.

1.5.2. Аппроксимированные зависимости

В ряде случаев зависимость $Y = f(X)$ известна (например, в результате ранее проведенных исследований), однако из-за сложности восприятия или по иным причинам не отвечает в полной мере поставленным целям. В подобных ситуациях функцию $f(X)$ нередко сознательно заменяют, *аппроксимируют* какой-то другой функцией $g(X)$, чаще всего – параметрической. Аппроксимированная зависимость, строго говоря, уже не отражает влияния X и Y , или, точнее, отражает ее искаженно. При этом близость f и g в какой-либо метрике не гарантирует от ошибок (например, близкие в равномерной метрике функции могут сильно отличаться значениями точек экстремума или перегиба). Поэтому замена f на g допустима только, если функция g обладает нужными для аппроксимации свойствами и определенным образом согласована с имеющейся исходной информацией. Так, если точно известно, что $Y = Y_i$ при $X = X_i$, то при аппроксимации должно соблюдаться равенство $Y_i = g(X_i)$. Сейчас аппроксимация зависимостей используется в основном при теоретических исследованиях (скажем, замена e^x на $1 + x$ при малых x стала здесь уже обычной), поскольку с широким внедрением компьютеров необходимость упрощения расчетных формул становится все менее актуальной. В этой связи сфера практического применения аппроксимированных зависимостей представляется нам ограниченной и мы не будем далее их рассматривать.

1.5.3. Нормализованные и регрессионные зависимости

При отыскании зависимости $Y = f(X_1, \dots, X_n)$ обычно неявно принимается, что характеристика объекта Y зависит *только* от характеристик (факторов) X_1, \dots, X_n . Однако ввиду предположения о целостности изучаемых объектов следует считать, что *все характеристики экономических объектов взаимосвязаны*, и потому априори на любой экономический показатель влияет обычно гораздо больше факторов, чем может быть учтено в математических моделях. Таким

образом, возникает необходимость разделения полного множества факторов на "существенные" X_1, \dots, X_n , прямо учитываемые в модели, и "несущественные", "прочие" – X_{n+1}, X_{n+2}, \dots , влияние которых приходится игнорировать (если среди них окажется фактор, достаточно сильно влияющий на Y , можно говорить о неправильной *спецификации* модели). Но даже если влияние каждого из "прочих" факторов мало, *совокупное* их влияние на Y может быть значительным – ошибка спецификации при этом лишь снижается, но не исчезает. Как же трактовать в этом случае установленную любым способом зависимость $Y = f(X_1, \dots, X_n)$? Обычно говорят, что такая зависимость не "точная", а приближенная. На наш взгляд, такие зависимости правильнее было бы называть *неполными*, ибо они учитывают *не полную*, а *сокращенную*, ограниченную совокупность влияющих на Y факторов. На практике используются два типа неполных зависимостей: нормализованные и регрессионные.

Нормализованные зависимости отражают влияние "существенных" факторов на Y "прочих равных условиях", т.е. при стабильных значениях "прочих факторов". Но к каким именно значениям "прочих факторов" будут относиться полученные результаты? При использовании статистических методов обычно постулируется, что такими значениями должны быть "средние" в том или ином смысле этого слова. Таким образом, объясняемой переменной становится не истинный (он же – наблюдаемый) показатель, а "идеальный" – тот, который мог бы быть достигнут при данных значениях учитываемых факторов и *средних* (или "*нормальных*") значениях прочих. Получаемая при этом зависимость "идеального" Y от "реальных" X_1, \dots, X_n может быть названа "нормализованной". Конечно, для ее практического построения хотелось бы и наблюдать эти "идеальные" значения Y . Однако такой возможности нет, и исследователь вынужден заменить измерение индикацией, "очищая" наблюдаемые значения Y от влияния "прочих факторов" (сглаживание, ввод корректирующих коэффициентов и т.п.; см. также п. 4.5). При этом возникает серьезная проблема применения построенной зависимости, например, когда при прогнозе значения Y будет сделана оговорка "при прочих средних условиях", строго говоря, обесценивающая весь прогноз в целом (поскольку прогнозные значения "прочих факторов" неизвестны и не обязаны совпадать со средними).

Мы видим, что искомую зависимость можно рассматривать как "обычную", такую зависимость $Y' = f(X_1, \dots, X_n)$, в которой наблюдения Y "идеального" показателя Y' содержат дополнительную "ошибку корректировки". Подчеркнем еще раз, что этот вывод относится к той, обычно имеющей место на практике ситуации, когда влиянием совокупности прочих факторов на Y нельзя пренебречь.

В общем случае "нормальные" значения прочих факторов могут отличаться от средних. Так, при анализе зависимости выпуска продукции от параметров технологического оборудования естественно исключить влияние возможных нарушений в технологии и организации

производства. Поэтому в построенной зависимости идеальный показатель Y' должен отразить объем выпуска при "нормальной", а отнюдь не "средней" технологии и организации производства, что предполагает использование специальных процедур "нормализации" ("приведения к нормальным условиям", "очистки от влияния возможных нарушений технологии и организации производства") наблюдаемых значений объемов выпуска.

Еще один способ "нормализации" приводит нас к *регрессионным* зависимостям.

Регрессионные зависимости. В ряде случаев влияние "прочих факторов" на объясняемую переменную формализуется путем ее представления как неопределенной величины. При этом вся совокупность "прочих факторов" X_{n+1}, X_{n+2}, \dots агрегируется в один неопределенный фактор ξ ("ошибку наблюдения"), а задача сводится к установлению функции $Y = f(X_1, \dots, X_n, \xi)$ или входящих в нее параметров от наблюдаемых переменных X_i и ненаблюдаемой переменной ξ . Естественно, что при этом приходится вводить дополнительные предположения не только о виде зависимости Y от ξ , но и о характере неопределенности ξ . Наиболее распространенной является вероятностная формализация, когда неопределенный фактор ξ трактуется как случайная величина (СВ) с неким стандартным (обычно – нормальным) распределением. Чаще всего здесь используется⁶ [1, 2 и др.] *модель аддитивной нормальной ошибки*: $Y = f(X_1, \dots, X_n) + \xi$, где ξ имеет нормальное распределение с нулевым средним и дисперсией $\sigma^2(X_1, \dots, X_n)$, которая в общем случае неизвестна и потому подлежит оценке наравне с "систематической" составляющей $f(X_1, \dots, X_n)$: последняя может трактоваться как математическое ожидание Y при заданных X_1, \dots, X_n . В "предельном" случае $n = 0$ модель $Y = a + \xi$ отвечает хорошо изученной задаче *определения среднего значения* (центра распределения). Несколько реже применяется аналогичная *модель мультипликативной нормальной ошибки* $Y = f(X_1, \dots, X_n)e^\xi$, которая сводится к предыдущей логарифмированием. Заметим, что в обоих случаях функции $f(X_1, \dots, X_n)$ можно придать иной и в чем-то более простой смысл – это значение Y , отвечающее нулевым (и в то же время средним) значениям неопределенного фактора ξ . Иными словами, функция $f(X_1, \dots, X_n)$ отражает здесь значение Y , "приведенное" к "нормальным" условиям, отвечающим $\xi = 0$. С этих позиций указанные типы регрессионных зависимостей можно одновременно считать и нормализованными.

В то же время нормализация путем приведения к случаю $\xi = 0$ может иногда вступать в противоречие с экономическим содержанием рассматриваемой характеристики объекта. Пусть, например, иссле-

⁶ При стандартной математико-статистической постановке задачи случайный фактор ξ не выделяется, а Y рассматривается как СВ, распределение которой зависит от X_1, \dots, X_n .

дуются зависимость продолжительности монтажных работ при строительстве небольших промышленных объектов от объемов этих работ. На подобном строительстве можно использовать и обычно используют башенные или самоходные стреловые краны. При "обычной" нормализации следовало бы выбрать какой-то тип крана (скажем, наиболее эффективный или наиболее распространенный) и "привести" к нему все наблюдаемые сроки выполнения работ. Если же этого не делать, отнеся тип крана к числу случайных факторов, полученная зависимость будет отвечать какому-то среднему между башенными и стреловыми типу крана, не существующему в природе, либо одновременно использованию тех и других кранов на одном объекте, что нерационально и обычно не делается.

Изложенная трактовка регрессионной зависимости несколько шире общепринятой: в статистике под ней понимают "только одну часть" модели аддитивной случайной ошибки, отражающую зависимость математического ожидания Y от X_1, \dots, X_n . При нашем же подходе составной частью регрессионной зависимости становится и функция $\sigma(X_1, \dots, X_n)$, отражающая "разброс ошибок", и, кроме того, допускаются иные виды неопределенности, не описываемые в терминах случайных величин (это обстоятельство будет рассмотрено в п. 3.3). Мы видим, что регрессионные зависимости вполне укладываются в общепринятую трактовку функциональных зависимостей, если только считать, что объясняемой переменной здесь является не истинное, а определенным образом нормализованное (приведенное к $\xi = 0$) значение объясняемой характеристики.

1.5.4. Структурные зависимости

Выше отмечалось, что значение $f(X)$ в регрессионной зависимости $Y = f(X) + \xi$ отражает определенным образом нормализованное значение Y . Естественным, что это значение неизмеримо, а его оценка может осуществляться только на основе наблюдаемых значений X и Y . В этой связи удобно разделить и обозначить по-разному измеряемые и неизмеримые переменные модели, после чего она примет следующий вид.

Каждый объект ГС характеризуется векторной измеримой характеристикой X и скалярной неизмеримой характеристикой V , о которой известно, что она зависит от X и *только* от X . Требуется оценить эту зависимость, исходя из результатов (X_i, Y_i) наблюдений некоторых объектов ГС. При этом наблюдаемые значения Y_i являются измеренными с ошибкой значениями V_i , причем характер ошибки описывается моделью: $Y_i = V_i + \xi_i$. При такой трактовке неизмеримой является только одна характеристика V , тогда как X предполагается измеримой. Между тем возможность наблюдать точные значения исследуемых характеристик имеется отнюдь не всегда. Гораздо чаще встречается ситуация, когда наблюдаемые показатели не вполне адекватно отражают нужные характеристики. При этом возникает более сложный тип зави-

симости – *структурная* зависимость. Типичной здесь является следующая постановка задачи.

Каждый объект ГС характеризуется двумя векторными и одной скалярной характеристиками X , U и V . Характеристика X измерима, т.е. наблюдается точно, тогда как U и V неизмеримы, т.е. результаты Y и Z их наблюдения содержат ошибки, например, случайные: $Y_i = U_i + A(X_i, U_i) \xi_i$, $Z_i = V_i + b(X_i, U_i) \bullet \xi_i$, где A – некоторая матрица, b – вектор, ξ_i – вектор с независимыми компонентами, имеющими стандартное нормальное распределение, \bullet – знак скалярного произведения. В этой ситуации необходимо оценить зависимость $V = f(X, U)$ по результатам наблюдений X_i , Y_i и Z_i . Естественно, что промежуточным этапом в решении такой задачи будет и оценка неизвестных зависимостей $A(X_i, U_i)$ и $b(X_i, U_i)$.

Идея замены множества "прочих факторов" одним "фактором неопределенности", положенная в основу регрессионного анализа, допускает развитие в различных направлениях. В частности, в подобной ситуации не обязательно ограничиваться анализом зависимости математического ожидания объясняемой переменной Y от объясняющих, можно исследовать и аналогичные зависимости для дисперсии Y или иных ее вероятностных характеристик. Развивая эту идею дальше, мы приходим к задаче установления зависимости закона вероятностного распределения объясняемой переменной Y от объясняющих переменных X_1, \dots, X_n . Такого рода зависимость также можно назвать "структурной", хотя этот термин используется в прикладной статистике для близких, но несколько иных моделей. Задачи оценки таких зависимостей намного сложнее, поскольку объектом оценки здесь являются вероятностные распределения.

По существу, рассматриваемые в статистике зависимости являются в большинстве своем структурными. Так, общая теория статистики рассматривает зависимость как "стохастическую (вероятностную) связь, при которой с изменением факторного признака меняется распределение единиц совокупности по результативному признаку" [17. С. 197]. На этом основании представляется, что структурные зависимости должны шире использоваться в экономических исследованиях.

2. ПРИНЦИПЫ ОПТИМАЛЬНОЙ ОЦЕНКИ ЗАВИСИМОСТЕЙ

Часто легче отстаивать принципы, чем жить по ним.

Эдлай Стивенсон

Обычно "качество" построенной статистической зависимости оценивается по тем или иным критериям. В конечном счете все эти критерии задаются априорно, однако между разными критериями есть принципиальные различия. Одни из них, как сумма квадратов отклонений, действительно, задаются априорно и непосредственно – это характерно для раздела прикладной статистики, известного под названием "анализ данных" [18, 19]. С другой стороны, такие критерии, как максимум правдоподобия, выводятся из существа рассматриваемых задач, хотя подобный вывод и нельзя считать математически строгим. В данном разделе мы рассмотрим различные критерии и покажем, что в целом второй, индуктивный путь построения критериев более плодотворен, а сфера его применения может быть существенно расширена.

2.1. Принцип минимального расстояния

2.1.1. Параметрический случай

Довольно часто теоретический анализ показывает, что искомые зависимости – параметрические, т.е. имеют вид *известной* функции $Y = \Phi(X, \mathbf{a})$ от (векторной) объясняющей переменной X и конечномерного вектора неизвестных параметров \mathbf{a} . Примером такой зависимости является модель Кейнса–Голдсмита зависимости сбережений S от дохода Y : $S = c(Y - b)$, где c – предельная склонность к сбережениям, b – доход, отвечающий нулевым сбережениям. Весьма интересный пример содержательного анализа, приводящего к параметрическому виду зависимостей, можно найти в теории производственных функций [4–6].

Простейшая задача оценки отвечает ситуации, когда ошибки наблюдений отсутствуют и Φ – полином от X с коэффициентами, образующими вектор \mathbf{a} (задача полиномиальной интерполяции). Ее решение дается интерполяционной формулой Ньютона. "Подгонка" параметров под результаты наблюдений издавна применялась в астрономии при описании движения планет. Со временем Птолемея предполагалось и

считалось законом, что планеты равномерно вращаются по окружностям вокруг Земли. Когда это пришло в противоречие с наблюдениями, закон уточнили – приняли, что планеты равномерно вращаются по окружностям (эпициклам) вокруг некоторых точек, которые, в свою очередь, равномерно вращаются по окружностям вокруг Земли. Задача при этом сводилась к отысканию радиусов указанных окружностей и скоростей вращения по ним на основе большого числа экспериментальных данных. Накопление астрономических наблюдений и повышение их точности, однако, показали, что модели такого рода расходятся с действительностью, в связи с чем было введено (между прочим, довольно естественное для эконометриков и инженеров) уточнение – предложили увеличить количество эпициклов. С математической точки зрения это были попытки приблизить формулу эллиптического (кеплеровского) движения планеты вокруг Солнца первыми членами ряда Фурье – тригонометрическим полиномом. Из этого примера видно, что неудачный выбор вида зависимости порой не только приводит к значительным вычислительным трудностям, но и способен на долгое время затормозить развитие науки.

Во многих случаях цель исследования состоит в оценке всех или наиболее важных параметров искомой зависимости. Так, рассчитав радиусы эпициклов и скорости обращения по ним, средневековый астроном мог спокойно "поставить точку" и сказать, что характер движения планет им установлен и дело остается лишь за тем, чтобы повысить точность найденных им величин, включив в расчеты последующие наблюдения.

Аналогичная ситуация имеет место и в экономических исследованиях, например, при оценке зависимости доходности акций данной фирмы от положения в экономике в целом, характеризуемого доходностью всего рыночного портфеля акций. Используемая здесь модель CAPM [20] учитывает специфику эмитента двумя параметрами, а ее точность определяется точностью оценки этих параметров. Однако здесь оцениваемых параметров несколько, и поиск оптимальных оценок становится многокритериальной задачей. С другой стороны, часто значимость разных параметров для исследователя не одинакова. Так, при построении производственной функции Кобба–Дугласа $\ln Z_t = \alpha \ln X_t + (1 - \alpha) \ln Y_t + \beta$ целью часто является оценка параметра α , отражающего эластичность национального дохода Z от капитальных ресурсов X , и "качество" решения определяется точностью оценки именно этого параметра.

Следует, однако, иметь в виду, что даже в задаче оценивания параметров в линейной регрессионной модели нельзя ответить на вопрос, какова же точность полученного решения (в указанном смысле). В лучшем случае здесь можно указать некоторую, например, 95%-ную доверительную область для вектора оцениваемых параметров, однако и в этом случае 5% исследователей, использующих данный метод, будут получать доверительные области, не содержащие истинного вектора параметров.

2.1.2. Непараметрический случай

При оценке непараметрической зависимости $Y = f(X)$ считаются известными некоторые свойства искомой функции, определяющие бесконечномерный класс функций \mathcal{F} в метрическом пространстве, которому она принадлежит. Но ни эти свойства, ни результаты наблюдений не могут дать полную и точную информацию о зависимости между характеристиками объекта. Поэтому трудно ожидать, чтобы функция f , согласованная пусть даже со всей имеющейся информацией, оказалась единственной – таких функций (будем называть их "допустимыми") будет достаточно много, и они образуют некоторое множество \mathcal{D} , в котором исследователь должен указать определенный (какой именно, зависит от применяемых методов оценки) элемент g . Если бы функция f стала известной, точность такого решения можно было бы оценить расстоянием $\rho(f, g)$ от f до g . Однако искомая функция f неизвестна. В этой связи точность решения задачи может оцениваться максимальным "расстоянием" $\rho(g, \mathcal{D}) = \sup_{f \in \mathcal{D}} \rho(g, f)$ между g и элементами \mathcal{D} , т.е. хаусдор-

фовым расстояние от g до \mathcal{D} ⁷. Отсюда вытекает постановка задачи оптимизации оценки, т.е. выбора такого решения g , которому отвечает минимальное расстояние $\rho(g, \mathcal{D})$. Тем самым оптимальная оценка здесь базируется на *принципе минимального расстояния*.

В ряде случаев конкретную задачу оценки зависимости можно рассмотреть как элемент некоторой последовательности таких задач с разным объемом исходной информации (так, задачу оценки параметров зависимости $Y = f(X)$ по известным значениям (Y_i, X_i) можно "вложить" в последовательность таких задач с возрастающим количеством узлов интерполяции X_i). Если точность оценки зависимости стремится к нулю с увеличением номера задачи, такую оценку (метод оценки) обычно называют *состоятельной*, а искомая зависимость становится *асимптотически измеримой*. Однако при малом количестве наблюдений сведения о состоятельности оценки могут принести исследователю скорее моральное удовлетворение. В подобной ситуации более информативной будет проверка используемого метода оценки на ряде стандартных тестовых задач, аналогично проверке эффективности численных алгоритмов оптимизации [21].

Применение принципа минимального расстояния оказалось плодотворным в приближенном анализе. Существенную роль здесь сыграли идеи А.Н. Колмогорова [22] и основанные на них оптимальные методы численного интегрирования, предложенные в [23]. Оказалось, что те же идеи могут быть использованы для решения задачи интерполяции – первые результаты такого рода получены в [24–27]. Здесь предполагается, что множество допустимых функций – компакт \mathcal{D} в не-

⁷ Разумно ввести "среднее расстояние" в функциональных пространствах не удастся, ибо там нет подходящей вероятностной меры, на которой можно было бы базировать исчисление подобных средних.

котором метрическом пространстве, метрика в котором характеризует "точность" (ошибку) интерполяции. Тогда имеющаяся информация – в данном случае ограничения $Y_t = f(X_t)$, $t = 1, \dots, T$, – сужает множество допустимых (согласованных с исходной информацией) функций до \mathcal{B} – некоторого компактного подмножества \mathcal{D} . Тогда в качестве искомой функции f можно взять центр шара минимального радиуса, описанного вокруг \mathcal{B} – при этом радиус шара будет характеристикой возможной ошибки. Отсюда естественно вытекает постановка задачи об *оптимальном выборе "узлов интерполяции"* X_t , обеспечивающих наибольшую точность интерполяции любых функций из \mathcal{D} (ищется минимальное R такое, что при любых Y_t все точки из \mathcal{D} , для которых $f(X_t) = Y_t$, содержатся в некотором шаре радиуса R).

В литературе, начиная с [21, 24, 25, 26], исследован широкий класс подобных задач. В частности, рассмотрены функции как одного, так и нескольких переменных, в том числе и комплексных, различные типы множеств допустимых функций и метрики (в частности, C и L_2). При этом построено много оптимальных или близких к ним интерполяционных формул и показана их состоятельность (с увеличением числа узлов интерполяции при не очень плохом их выборе ошибка оценки стремится к нулю). Установлена и связь оптимизационных задач теории оценки функций с поперечниками множеств в функциональных пространствах [24, 28]. Однако все результаты такого рода получены при весьма специфических ограничениях – предполагаются известными пределы изменения производных искомой функции (ограничениям на производные более высокого порядка отвечают более точные интерполяционные формулы). Эти результаты нашли применение в задачах табулирования и при разработке стандартных программ для ЭВМ [21], однако их применимость в экономике сомнительна, ибо исходная информация здесь относится к относительно небольшому числу объектов и не всегда позволяет определить границы производных искомой функции. Наличие ограничений на "гладкость" функции здесь принципиально. Если неизвестно, насколько гладкой является функция, сколь велики могут быть ее производные, какова ее "аналитическая структура", класс допустимых функций предельно расширяется, а ошибка решения, например, за счет неточного определения точек разрыва или резкого роста, может оказаться какой угодно⁸.

В то же время данный подход применим и при наличии ошибок наблюдения. Так, если вместо истинных значений Y_t и X_t наблюдаются значения U_t и V_t и ошибки этих наблюдений не превосходят соответственно ε и δ , то на множество \mathcal{D} допустимых функций накладываются дополнительные ограничения: $|X_t - U_t| \leq \delta$, $|f(X_t) - V_t| \leq \varepsilon$. Однако, если ошибки – случайные, такой подход уже неприменим (п. 2.7).

⁸ Например, если ищется зависимость затрат на изготовление машины от какого-либо технического параметра, то обычно априори неизвестно, не возникают ли резкие изменения затрат в каком-либо диапазоне значений этого параметра.

Интересные ситуации возникают, когда класс "допустимых" функций не является компактом. Пусть, например, Y – *неубывающая* функция от X и $A = X_1 < \dots < X_T = B$. Очевидно, что при $X_t \leq x \leq X_{t+1}$ будет $Y_t \leq f(x) \leq Y_{t+1}$, причем эти границы неуплучшаемы. Поэтому наименьшая (в равномерной метрике) ошибка интерполяции достигается, если взять в качестве $f(x)$ монотонную разрывную функцию, принимающую значения Y_t в точках X_t и $(Y_t + Y_{t+1})/2$ – в интервалах $X_t < x < X_{t+1}$. В то же время указать заранее, сколько узлов интерполяции потребуется для достижения заданной точности, невозможно, как нельзя указать и какие-либо *двусторонние* границы для искомой функции вне "интервала наблюдений" $[A, B]$ ⁹. Тот же подход можно использовать, если известно, что искомая функция *выпукла или вогнута*. Скажем, для вогнутой $f(x)$ на отрезке $X_t \leq x \leq X_{t+1}$ при $1 < t < T - 1$ имеют место неравенства

$$\begin{aligned} f^-(x) &= Y_t + \frac{Y_{t+1} - Y_t}{X_{t+1} - X_t}(x - X_t) \leq f(x) \leq f^+(x) = \\ &= \min \left\{ Y_t + \frac{Y_{t-1} - Y_t}{X_{t-1} - X_t}(x - X_t); Y_{t+1} + \frac{Y_{t+2} - Y_{t+1}}{X_{t+2} - X_{t+1}}(x - X_{t+1}) \right\} \end{aligned}$$

(при $t = 1$ и $t = T - 1$ из двух величин, стоящих здесь под знаком минимума, остается только одна). Поэтому наилучшей оценкой для $f(x)$ будет вогнутая функция, равная $(f^+(x) + f^-(x))/2$ на отрезке (X_t, X_{t+1}) . Разбивая отрезок $[A, B]$ узлами интерполяции на все большее число равных частей, можно получить, что любая вогнутая на $[A, B]$ (но не на всей числовой оси!) функция будет асимптотически измеримой.

Данный подход (он применим и в многомерном случае, однако расчетные формулы становятся сложнее) явно ориентирован не на эконометрические, а на технические или вычислительные задачи, отличающиеся характером исходной информации. Вероятно, он может эффективно использоваться в стандартизации и техническом нормировании. Однако в эконометрике сфера его применения ограничивается, по нашему мнению, зависимостями между экономическими и физическими или техническими показателями объектов. По-видимому, он применим при установлении зависимости затрат на добычу полезного ископаемого от глубины залегания или себестоимости изделия от серийности выпуска (подобных зависимостей на практике довольно много, но теоретических моделей для большинства из них не разработано). Однако здесь важно учесть, что такие традиционные для математики метрики, как равномерная или интегральная, для рассматриваемого класса задач обычно неудобны. Например, в равномерной метрике расстояние между функциями $Y = 2X$ и $Y = 2,01X$, если их рассматривать на всей положительной полуоси, бесконечно велико, в то время как любому эконо-

⁹ Такие границы можно указать в других ситуациях, например, когда известно, что $\ln f(x)$ – вогнутая функция, стремящаяся к $-\infty$ при $|x| \rightarrow \infty$.

мисту-практику такие функции покажутся близкими. Наоборот, в той же метрике зависимости $Y = X$ и $Y = X + 0,1$ близки, хотя в конкретной ситуации эти функции для экономиста могут показаться совершенно различными (например, в ситуации, когда X – планируемая мощность предприятия, Y – затраты на его строительство, то для экономиста очевидно, что на строительство завода нулевой мощности нужны нулевые затраты, и потому вторая зависимость должна быть отвергнута и не может считаться близкой к первой).

Таким образом, хотя подход, связанный с введением "традиционных математических" расстояний между зависимостями и является формально корректным, его практическое применение требует большой осторожности. При введении метрики в пространстве функций необходимо учитывать и цели исследования. Например, в [6, 29] рассмотрена ситуация, когда целью исследования является возможно более точное определение "поверхности уровня" искомой зависимости, т.е. множества таких значений аргумента X , которым отвечает данное значение Y . Отсюда вытекает целесообразность измерения расстояния между функциями через хаусдорфово расстояние между их "поверхностями уровня" (см. также п. 2.3).

2.2. Принцип минимальных потерь

Часто конечной целью построения зависимостей между характеристиками экономических объектов является исследование и оптимизация хозяйственных решений. В этом случае ошибка при оценке зависимости сопряжена с возможными финансовыми или иными потерями (ущербом), которые хотелось бы минимизировать. Это означает, что в основу получения оптимальной оценки зависимости может быть положен принцип *минимизации потерь $\pi(f, g)$ от применения g вместо f* . Соответственно точность решения g будет характеризоваться максимумом величины $\pi(f, g)$ по всем допустимым функциям f , а оптимальным решением будет функция g , минимизирующая эти максимальные потери. Именно такой подход использован, например в [1], при описании построения регрессионной функции как решения задачи оптимизации. Отметим, однако, что функция потерь $\pi(f, g)$ не обязательно обладает свойствами расстояния (например, может не выполняться неравенство треугольника), так что рассматриваемый принцип не совпадает с принципом минимального расстояния.

Опираясь на изложенный подход и используя стоимостные или натуральные оценки потерь, можно узнать, насколько вырастут потери пользователя искомой зависимости (*заказчика*), если он будет опираться на другую, неточную зависимость. Это, безусловно, является преимуществом данного подхода. В то же время присущий ему недостаток есть продолжение его достоинств: для построения экономически интерпретируемой функции потерь необходимы специальные экономические

исследования (включая и экономико-математическое моделирование потерь) – это существенно повышает требования к исходной информации. К тому же величина потерь всегда зависит от цен и других быстроменяющихся характеристик экономического окружения, что вынуждает пересматривать оценки зависимостей при изменении внешней ситуации.

2.3. Принцип минимума интегрального отклонения

Точность оценки неизвестных зависимостей нередко оценивается следующим образом. Пусть истинные значения показателей Y и X связаны зависимостью $Y = f(X)$, а по наблюдаемым значениям (Y_i, X_i) оценена (построена) зависимость $Y = \Phi(X)$. При этом равенства $Y_i = \Phi(X_i)$ могут не выполняться. Точность восстановления истинного значения $\Phi(X_i)$ ("локальная" точность найденной зависимости) может быть охарактеризована показателем (*локального*) отклонения – некоторой функцией $u_i = \phi(X_i, Y_i)$, обращающейся в нуль при $Y_i = \Phi(X_i)$ ¹⁰.

Вопрос о рациональном выборе функций ϕ нетривиален и заслуживает более подробного рассмотрения. Обычно в этом качестве выступают абсолютное отклонение $u_i = |Y_i - \Phi(X_i)|$ и относительные отклонения $u_i = |Y_i/\Phi(X_i) - 1|$ и $u_i = |\Phi(X_i)/Y_i - 1|$. Однако специфика экономических исследований обуславливает применение в некоторых случаях и иных, недостаточно хорошо исследованных в математическом аспекте методов измерения отклонений (мы не останавливаемся здесь на важном вопросе, в каких именно ситуациях целесообразно использовать тот или иной метод). Так, использование абсолютных и относительных отклонений оправдано, когда объясняемая характеристика Y наблюдается с ошибками, однако это перестает быть естественным, если ошибки присущи наблюдениям объясняющих переменных. В подобных ситуациях отклонение u_i полезно измерять расстоянием от точки X_i до изокванты $\text{Isoq}(Y_i) = \{X | \Phi(X) = Y_i\}$ [6]. Однако расстояние от точки до множества может вводиться по-разному. Так, u_i можно определить как евклидово расстояние от точки X_i до ближайшей к ней точки $X_i' \in \text{Isoq}(Y_i)$. Этот способ не учитывает "относительной весо-мости" отдельных характеристик (компонент вектора X), однако указанный недостаток может быть устранен, если эти характеристики предварительно нормировать или, что то же, использовать для определения расстояния не простую, а взвешенную сумму квадратов разностей $(X_{ii} - X'_{ii})$. Между тем можно учесть, что экономическое содержание переменных X позволяет ввести в пространстве характеристик дополнительное отношение "сходства". Так, если компоненты вектора X отражают объемы ресурсов, производимых или затрачиваемых наблю-

¹⁰ Для регрессионных зависимостей отклонения u_i часто трактуются как "потери от неточности восстановления значения" математического ожидания Y при $X = X_i$.

даемым объектом, возникают два естественных отношения "сходства" ("общности"):

1) сходными с вектором X считаются вектора X' вида λX , имеющие ту же *структуру*, то же соотношение между ресурсами;

2) сходными с вектором X считаются вектора X' , имеющие ту же *ценность*, т.е. стоимостную оценку ресурсов. Если обозначить через p вектор цен на ресурсы, то условием "сходства" будет равенство $p \bullet X' = p \bullet X$, где \bullet – знак скалярного произведения.

В подобных ситуациях более полный учет "качественной" информации об объектах ГС и экономическом содержании исследуемых характеристик этих объектов может быть обеспечен, если определить отклонение u_i как евклидово расстояние (при необходимости – с весами)

от точки X_i до ближайшей к ней *сходной* точки изокванты $X'_i \in \text{Isoq}(Y_i)$.

Вычисление таких отклонений часто не составляет особой сложности. Пусть, например, функция $\Phi(X)$ – однородная степени h по совокупности входящих в нее переменных, а сходными считаются пропорциональные вектора. Тогда единственной точкой изокванты $\text{Isoq}(Y_i)$, "сходной" с

точкой X_i , будет $X'_i = \lambda_i X_i$, где $\lambda_i = \left[\frac{Y_i}{\Phi(X_i)} \right]^{1/h}$. Этому отвечает

относительное отклонение $|\lambda_i - 1|$ и абсолютное отклонение $u_i = \|X'_i - X_i\| = \|X_i\| \cdot |\lambda_i - 1|$, где $\|X_i\|$ – расстояние от точки X_i до начала координат.

При правильном измерении локальных отклонений они, безусловно, характеризуют точность соответствующей зависимости. Так, если все они малы, соответствующую зависимость обычно считают достаточно точной. Однако чаще имеет место ситуация, когда для одних i отклонения u_i малы, для других – велики. В этой связи необходимо ввести дополнительную характеристику – *интегральное отклонение*, агрегирующую информацию о величине отдельных локальных отклонений. Ее естественно представлять как некоторую функцию $Q(u_1, \dots, u_T)$ от независимых локальных отклонений u_1, \dots, u_T и использовать в качестве критерия оптимального оценивания (отметим, что в качестве критериев оптимальной оценки зависимостей функция Q и любая монотонная функция от Q эквивалентны).

Агрегирующая функция Q при данном методе обычно задается априорно, однако разумные способы агрегирования локальных отклонений (или "потерь") обязательно должны удовлетворять двум условиям "нормальности":

- если все локальные отклонения равны нулю, то и интегральное отклонение тоже равно нулю: $Q(0, \dots, 0) = 0$;

- при увеличении любого локального отклонения интегральное отклонение возрастает.

Иногда от агрегирующей функции требуется еще и сим-

метричность: интегральное отклонение должно зависеть лишь от величин локальных отклонений, но не от того, в каком порядке они занумерованы. Это требование, однако, может быть неоправданным применительно к обработке временных рядов или, в более общем случае, несинхронных наблюдений экономических объектов.

Обычно для агрегирования применяются такие нормальные функции, как: сумма квадратов отклонений, сумма модулей, максимальный из модулей отклонений и т.д. Выбор между ними или обращение к другим функциям неочевидны. В этой связи представляется целесообразным использование аксиоматического подхода к построению показателя интегрального отклонения, когда вид агрегирующей функции выводится из требуемых ее свойств "дифференциального" характера. Для их

формулировки введем показатели $q_t(u_1, \dots, u_T) = \frac{\partial Q(u_1, \dots, u_T)}{\partial u_t}$,

($t = 1, \dots, T$), выражающие "чувствительности" интегрального отклонения по отношению к изменению локальных. Следующее утверждение [15] (близкие результаты, выраженные в терминах эластичностей замещения одних факторов другими, доказаны в [30, 31, 32]) дает аксиоматическое описание широкого класса агрегирующих нормальных функций.

Утверждение. Пусть количество наблюдений больше 2. Если функция Q нормальная и чувствительность интегрального отклонения по отношению к изменению каждого из локальных зависит только от доли последнего в интегральном отклонении, то

$$Q = [(b_1 u_1)^\beta + \dots + (b_T u_T)^\beta]^{1/\beta}, \quad (2.1)$$

где $\beta \neq 0$, b_1, \dots, b_T – положительные константы.

Доказательство. Тот факт, что величина q_t зависит только от отношения u_t/Q , может быть выражен соотношениями:

$$\frac{\partial Q(u_1, \dots, u_T)}{\partial u_t} = f_t \left(\frac{u_t}{Q} \right), \quad (t = 1, \dots, T), \quad (2.2)$$

где f_t – некоторые неизвестные функции одного переменного.

Дифференцируя (2.2) по другой переменной u_s , получим:

$$\frac{\partial^2 Q}{\partial u_s \partial u_t} = \frac{\partial}{\partial u_s} f_t \left(\frac{u_t}{Q} \right) = -\frac{u_t}{Q^2} f_t' \left(\frac{u_t}{Q} \right) \frac{\partial Q}{\partial u_s} = -\frac{u_t}{Q^2} f_t' \left(\frac{u_t}{Q} \right) f_s \left(\frac{u_s}{Q} \right).$$

Это равенство должно сохраниться, если номера наблюдений s и t поменять местами, поэтому $-\frac{u_t}{Q^2} f_t' \left(\frac{u_t}{Q} \right) f_s \left(\frac{u_s}{Q} \right) = -\frac{u_s}{Q^2} f_s' \left(\frac{u_s}{Q} \right) f_t \left(\frac{u_t}{Q} \right)$, откуда следует, что

$$H_t(u_t / Q) = \frac{u_t f_t'(u_t / Q)}{Q f_t(u_t / Q)} = \frac{u_s f_s'(u_s / Q)}{Q f_s(u_s / Q)} = H_s(u_s / Q).$$

Поскольку $T > 2$, возьмем какое либо $k \in [1, T]$, отличное от s и t . Меняя подходящим образом u_k , можно добиться, чтобы при изменении u_s и u_t величина Q не менялась. Поэтому при всех возможных значениях аргумента функция H_t , а значит и H_s , будет принимать одно и то же постоянное значение, которое мы обозначим через $\beta - 1$. Поскольку номера наблюдений s и t могут быть выбраны произвольно, следует, что $H_s(z) = \frac{zf'_s(z)}{f_s(z)} = \beta - 1$ при всех s . Решая это уравнение, найдем $f_s(z) = a_s z^{\beta-1}$, где a_s – некоторая константа. Но тогда в силу (2.2) имеем:

$$\frac{\partial Q}{\partial u_t} = a_t \left(\frac{u_t}{Q} \right)^{\beta-1}, \quad (t = 1, \dots, T). \quad (2.3)$$

Полученная система дифференциальных уравнений имеет два типа решений.

1) если $\beta = 0$, то $Q = u_1^{a_1} \dots u_T^{a_T}$. Такая функция не нормальна: если $u_1 = 0$, то при $a_1, \dots, a_T > 0$ и любых значениях других u_s интегральное отклонение будет оставаться равным нулю;

2) если $\beta \neq 0$, то $Q = (a_1 u_1^\beta + \dots + a_T u_T^\beta + b)^{1/\beta}$. Эта функция будет нормальной только, если $b = 0$, а все a_t – положительны. Но тогда она имеет вид (2.1). ■

Потребовав дополнительно, чтобы функция Q была симметричной, из доказанного утверждения легко получить, что (с точностью до постоянного множителя)

$$Q = (u_1^\beta + \dots + u_T^\beta)^{1/\beta}. \quad (2.4)$$

При $\beta = 1$ или 2 полученный критерий приводит к методам наименьших модулей или наименьших квадратов. В первом случае чувствительность интегрального отклонения по отношению к изменению каждого из u_t – постоянная величина, во втором – пропорциональна доле u_t в интегральном отклонении. При этом в обоих случаях могут использоваться не только абсолютные, но относительные отклонения (это предлагалось в [6, 33]) или отклонения от изокванты.

Предельными для функций (2.1) при $\beta \rightarrow \infty$, являются функции вида $Q = \max\{b_1 u_1, \dots, b_T u_T\}$, из которых симметричной будет только одна (с точностью до постоянного множителя): $Q = \max\{u_1, \dots, u_T\}$. Эти функции не являются нормальными: если, например, отклонение u_1 мало по сравнению с отклонениями других наблюдений, его увеличение не сказывается на величине интегрального отклонения Q . Тем не менее в ряде случаев такие агрегирующие функции также используются в эконометрических расчетах (последняя отвечает минимаксному методу оценки зависимостей).

Использование показателя интегрального отклонения как критерия оптимальной оценки зависимостей широко распространено и часто дает практически полезные результаты, однако такой метод нельзя считать универсальным по следующим причинам.

1. Способ измерения локальных отклонений и вид агрегирующей функции Q , как правило, задаются *экзогенно*, а не выводятся логически из имеющейся информации. В результате Q не может трактоваться как оценка совокупных потерь от неточного восстановления функции, что предложено в [1. С. 168] (так, без специального экономического расчета трудно представить, чтобы в реальной ситуации подобные потери были пропорциональны сумме квадратов локальных отклонений или хоть как-то скоррелированы с ней).

2. Минимизация Q не позволяет учесть информацию о степени влияния "прочих факторов". Так, трудно считать достаточно точной рассчитанную этим методом зависимость урожайности от количества вносимых удобрений, если известно, что она во многом определяется погодными условиями. Аналогично нельзя считать достаточно точной зависимость, обеспечивающую среднеквадратичное отклонение 2, если мы знаем, что случайная ошибка наблюдения имеет дисперсию не менее 100.

3. При построении агрегирующей функции ее аргументы (u_i) предполагаются независимыми. Поэтому критерий интегрального отклонения не учитывает имеющуюся информацию о зависимости между ошибками наблюдений разных объектов (или одного и того же объекта в разные моменты времени), например, о том, что $u_1 = u_2$ или $u_1 \geq u_2 + u_3$.

При оценке параметрических зависимостей методом минимизации интегрального отклонения обычно не удается правильно учесть часто встречающиеся в экономике ограничения на вид оцениваемой зависимости (например, обеспечить монотонность $\Phi(X)$ по переменной X_1 и ее выпуклость по переменным X_2 и X_3). Пытаясь преодолеть эту трудность, на практике нередко варьируют составом объясняющих переменных и количеством вводимых параметров, заменяя содержательный анализ математическими спекуляциями. В этой связи прикладная статистика изучает проблему обеспечения "компромисса между степенью общности привлекаемого класса допустимых решений и точностью оценивания, которой возможно при этом добиться", и ставит вопросы типа "нужно ли для достижения максимальной точности в задаче восстановления значений результирующего показателя по значениям предикторов включать в модель *все* предикторные переменные, а если не все, то *сколько и какие именно*?" [1. С. 190–191]. Вопросы такого рода оправданы, по нашему мнению, только в случаях, когда речь идет об установлении каких-либо зависимостей лицами, не являющимися специалистами в соответствующей области, которые могут по своему произволу менять состав учитываемых факторов. В квалифицированных же исследованиях как состав объясняющих показателей, так и вид зависимости устанавливаются исходя из существа исследуемой проблемы.

Таким образом, метод минимизации интегральных отклонений отнюдь не универсален, а точность оцененной этим способом зависимости не всегда характеризуется величиной отклонений.

2.4. Принцип максимальной гладкости в задаче интерполяции

Как отмечалось в п. 2.1.2, решение задачи интерполяции (в нашей терминологии – оценки) можно получить, используя принцип минимального расстояния. Для этого, однако, необходимы экзогенно заданная метрика в пространстве функций и компактность множества допустимых функций в этой метрике. К тому же построение оптимальных интерполяционных формул обычно представляет трудную математическую проблему. В этой связи представляет интерес рассмотреть более простой подход к решению задачи, имеющий и более широкую сферу применения.

Рассмотрим задачу интерполяции функции одного переменного $f(X)$ по ее значениям в некоторых точках. При этом известно, что $f'(X)$ непрерывна, однако пределы ее изменения не известны. Очевидно, что решением такой задачи могут быть любые гладкие функции, график которых проходит через заданные точки. Среди них, естественно, могут быть и сколь угодно сильно колеблющиеся. Так, если наблюдения производились при целых X , то вместе с $f(X)$ *полностью* согласованной с исходной информацией будет и любая функция вида $f(X) + A \sin(2\pi kX)$. Однако на практике подобные функции отвергают, считая их "неестественными", изменяющимися "недостаточно плавно". Это позволяет выдвинуть гипотезу о том, что существует некоторый измеритель "гладкости" ("плавности"), по которому исследователь отличает "хорошие" ("гладкие") функции от "плохих" и отбирает лучшую из них.

Что может служить измерителем гладкости? Очевидно, какой-то интегральный показатель величин производных искомой функции. Оказывается, что если в качестве измерителя гладкости используется функционал $Q_1 = \int |f'(X)|^2 dX$, то наиболее "гладкой" (минимизирующей этот функционал) функцией окажется кусочно-линейная, получающаяся при соединении отрезками соседних точек ("обычная" линейная интерполяция). Однако такое решение часто не удовлетворяет исследователей. Поэтому конструктора решают подобную задачу, соединяя соответствующие точки обычно тонкой линейкой, подпирая ее упорами ("криками") в этих точках с той или другой стороны¹¹. В результате получается сплайн [34, 35, 36], минимизирующий функционал $Q_2 = \int |f''(X)|^2 dX$ ¹². Обратим внимание, что интегралы здесь могут браться по всей числовой оси. При этом вне отрезка, на котором

¹¹ В США пользовались гибкими рейками (spline), к которым подвешиваются свинцовые грузила.

¹² Он построен путем "склеивания" многочленов третьего порядка, проходящих через данные точки графика.

сосредоточены наблюдаемые значения X , "оптимальная" функция окажется для Q_1 – константой, для Q_2 – линейной. Тем самым обеспечивается не только интерполяция, но и экстраполяция наблюдаемых значений функции (в многомерном случае это одно и то же: про три произвольные точки пространства уже нельзя сказать, что одна из них лежит между двумя другими).

Метод сплайнов применим и в случае, когда значения функции Y_t наблюдаются неточно. Так, если известно, что ошибки наблюдений не превосходят d , то задача сводится к минимизации функционала Q_2 при ограничениях $|Y_t - f(X_t)| \leq d, \forall t$.

Функционалы гладкости применимы и для оценки функций многих переменных. Так, в двумерном случае хорошие результаты дает применение функционала $Q_3 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\Delta f|^2 dx dy$, где Δ – знак оператора

Лапласа. Как и в одномерном случае, решение задачи здесь сведется к решению системы линейных уравнений, но "оптимальная" функция уже не будет кусочно-полиномиальной.

Более простые результаты можно получить в ситуации, когда известны какие-либо дополнительные свойства оцениваемой функции. Приведем два примера.

Пример 2.1. Пусть оценивается *сепарабельная* функция $Y = f_1(X_1) + \dots + f_n(X_n)$ по наблюдаемым значениям $(Y_t, X_{1t}, \dots, X_{nt})$. Здесь удобно использовать функционал $Q = \sum_{i=1}^n \int_{-\infty}^{\infty} |f_i''(X_i)|^2 dX_i$, и оптимальные

функции, минимизирующие этот функционал, оказываются одномерными сплайнами. Правда, решение здесь не единственно (к любой из функций f_i можно добавить произвольную константу, а из любой другой – вычесть эту константу), но на значение Y это не влияет, а для получения однозначного решения достаточно наложить, например, ограничение $f_2(0) = f_3(0) = \dots = f_n(0) = 0$. ■

Пример 2.2. Пусть требуется оценить *однородную* функцию известной степени α от двух переменных $Z = F(X, Y)$ по ее значениям в точках (X_t, Y_t) . Такая функция определяется своими значениями на единичной окружности. Перейдем поэтому к полярным координатам и обозначим: $X_t = R_t \cos \varphi_t$; $Y_t = R_t \sin \varphi_t$; $Z_t = R_t^\alpha U_t$; $U_t = f(\varphi_t)$. Теперь задача сводится к отысканию периодической функции f по ее значениям в точках φ_t , для чего можно использовать функционал

$Q_2 = \int_0^{2\pi} |f''(\varphi)|^2 d\varphi$. Оптимальным решением при этом будет периодический сплайн. ■

Функционал гладкости, однако, должен учитывать и иную исходную информацию об оцениваемой функции. Так, при интерполяции монотонной функции с помощью функционала Q_2 полученный сплайн может оказаться не монотонным. Здесь необходимо применять иные

функционалы, например,

$$Q_4 = \int \{f''(X) / f'(X)\}^2 dX, \quad (2.5)$$

что, однако, приводит к решению системы нелинейных уравнений.

Казалось бы, использование функционалов гладкости должно приводить к более субъективным решениям по сравнению с применением функций потерь. На деле этого не происходит, ибо сами функции потерь обычно задаются весьма приблизительно и с высокой долей субъективизма, в то время как представления о "степени гладкости" зависимостей у разных специалистов близки (в экспериментах установлено, например, что разные математики, графически решая одну и ту же задачу интерполяции, получают довольно близкие результаты). К тому же функционал гладкости во многих случаях также допускает трактовку как своеобразная функция потерь, так как выбор "негладкой" зависимости может усложнить производственный процесс и часто создает трудности во взаимоотношениях с заказчиком или оппонентами¹³. Поэтому использование функционалов гладкости для оценки непараметрических зависимостей в экономике может оказаться весьма полезным, и далее этот подход будет развит.

2.5. Принцип максимального правдоподобия

Каков бы ни был способ оценки неизвестной зависимости, исследователю приходится учитывать общее требование согласованности получаемой зависимости с имеющейся информацией. Но как понимать такую согласованность при наличии случайных ошибок? Общая теория статистики [17, 37] на этот вопрос ответа не дает, ссылаясь на "общепринятый" метод наименьших квадратов, который, как нам кажется, каждый раз нуждается в содержательном обосновании исходя из условий поставленной задачи. Один из путей такого обоснования иллюстрируется следующим примером.

Пример 2.3. Пусть оценивается линейная зависимость между переменными X и Y , причем X наблюдается точно, а Y – со стохастической ошибкой, имеющей нормальное распределение с небольшой дисперсией. Пусть имеются три наблюдения пары значений X и Y : $(0; -0,01)$, $(1; 1,02)$ и $(2; 1,99)$. Надо полагать, что любой метод статистической обработки покажет наличие зависимости, близкой к $Y = X$. Однако наблюдениям "не противоречит" и зависимость $Y = 4 - X$, поскольку теория вероятностей *не исключает* возникновения соответствующих такому выводу ошибок при измерении Y . Однако такую зависимость статистики отбрасывают: она могла бы иметь место только при очень больших и маловероятных ошибках наблюдения, и потому гораздо хуже согласуется с данными наблюдений, чем "рассчитанная" зависимость

¹³ В последнем случае исследователь должен ориентироваться не столько на "свой" функционал гладкости, сколько на функционал, которым пользуются его предполагаемые оппоненты.

$Y = X$. Формализуется это следующим образом. Вероятности событий, которые фактически произошли (хотя при данных условиях могли осуществиться и другие события), в теории вероятностей характеризуются термином "правдоподобие". В этих терминах зависимость $Y = 4 - X$ описывается как в принципе возможная, но *значительно менее правдоподобная* по сравнению с зависимостью $Y = X$. Тем самым подразумевается возможность, пусть экспертно, отделить "достаточно правдоподобные" зависимости от "практически неправдоподобных" и тем самым руководствоваться некоторым ограничением на величину "допустимого правдоподобия" (этим целям служат и критерии математической статистики, выделяющие область в пространстве оцениваемых параметров, где значение функции правдоподобия достаточно велико). Таким образом, здесь требование согласованности с имеющейся информацией трансформируется в требование "достаточно-го правдоподобия" искомой зависимости. ■

Рассмотрение соответствующих методов начнем с ситуации, когда случайный вектор X имеет дискретное распределение и вероятность события $X = Y$ – известная функция $P(Y, \Theta)$ от Y и неизвестного параметра Θ . Произведено наблюдение, показавшее, что $X = H$. Требуется оценить значение Θ .

Величина $P(H, \Theta)$ здесь показывает, с какой вероятностью повторение наблюдений будет приводить к тому же результату при данном Θ . Если она мала, то наблюденное значение H маловероятно и плохо согласуется с известным распределением. В [38] предложено, по существу, обратить это утверждение и считать, что *значение параметра Θ при малом $P(H, \Theta)$ плохо согласуется с результатом наблюдения, а при большом $P(H, \Theta)$ – хорошо с ним согласуется*. Для описания такой согласованности предложен термин "**правдоподобие**", так что $P(H, \Theta)$, рассматриваемая при фиксированном H как функция от Θ , именуется **функцией правдоподобия**. Соответственно оптимальной оценкой предлагается считать такой Θ , при котором $P(H, \Theta)$ *максимально*: в этом суть **принципа максимального правдоподобия** (МП-принципа) и основанного на нем метода максимального правдоподобия (МП-метода), широко применяемому в математической статистике со времен появления работы [38]. Чаше, однако, МП-принцип используется для недискретных распределений. Здесь вероятности событий типа $X = H$ обычно нулевые, однако можно говорить о вероятностях попадания случайного вектора X в малую окрестность той иной точки H , который обычно описываются с помощью плотности распределения $p(H, \Theta)$. Поэтому здесь функцией правдоподобия именуется уже **плотность $p(H, \Theta)$** , рассматриваемую при фиксированном H как функцию от Θ . Соответственно модифицируется и МП-метод [2]. Далее мы увидим, что подобный переход от вероятностей к плотностям иногда может создать серьезные трудности. Использование МП-метода позволяет учесть и "неравноценность" наблюдений (разную дисперсию ошибок разных наблюдений, гетероскедастичность).

Пример 2.4. Оценивается удельный расход ресурса (a) при производстве некоторой продукции. Для этого используются фактические данные (X_i, Y_i) об объемах выпуска продукции и потребления ресурса по годам (или месяцам) отчетного периода. Казалось бы, для описания соответствующей регрессионной зависимости можно было бы использовать модель $Y_i = aX_i + \xi_i$, где ξ_i — независимые нормально распределенные ошибки с нулевым средним и одинаковой дисперсией S . Здесь плотность совместного распределения Y_i имеет вид $\prod_i \frac{1}{\sqrt{2\pi S}} \exp\{-(Y_i - aX_i)^2 / 2S\}$. Максимум этого выражения по S

и a достигается при $a = \sum_i X_i Y_i / \sum_i X_i^2$; $S = \frac{1}{T} \sum_i (Y_i - aX_i)^2$ и равен $[2\pi eS]^{-T/2}$. В то же время более оправданным будет иное предположение. Расход ресурса определяется обычно суммированием данных о расходе отдельных его порций, причем измерения объема каждой порции производятся с независимыми ошибками. Естественно считать, что количество таких порций пропорционально объему производства. Поскольку при суммировании независимых ошибок их дисперсии суммируются, из этого следует, что дисперсия ξ_i пропорциональна X_i . Обозначив коэффициент пропорциональности через R , получим

новую функцию правдоподобия $\prod_i \frac{1}{\sqrt{2\pi R X_i}} \exp\left\{-\frac{(Y_i - aX_i)^2}{2R X_i}\right\}$. Ее максимальное значение $\left[(2\pi e R)^T \prod_i X_i\right]^{-1/2}$ достигается при

$$a = \sum_i Y_i / \sum_i X_i; \quad R = \frac{1}{T} \sum_i \frac{(Y_i - aX_i)^2}{X_i}.$$

В отличие от предыдущей, данная оценка a имеет более простой экономический смысл и обычно применяется на практике. ■

Как видно из примера, при оценке зависимости нужна не только теоретическая модель этой зависимости, но и теория, описывающая характер ошибок наблюдений.

Пример 2.5. Оценивается параметр a зависимости $Y = \Phi(X, a)$, связывающей истинные значения показателей некоторой ГС объектов, причем истинные значения X_i измеряются точно, а вместо истинных значений Y_i наблюдаются значения СВ V_i , имеющие нормальные распределения со средними Y_i и дисперсией S (неважно, известной или нет). Функция правдоподобия (плотность вероятностей всех наблюдаемых значений) здесь равна $\prod_i \frac{1}{\sqrt{2\pi S}} \exp\{-(V_i - \Phi(X_i, a))^2 / 2S\}$, а наиболее правдоподобный, максимизирующий эту функцию век-

тор параметров a может быть найден по критерию $\sum_i [V_i - \Phi(X_i, a)]^2 \Rightarrow \min$ (метод наименьших квадратов). Именно в такой модификации данный метод чаще всего применяется в экономических исследованиях. Если же в рассмотренной ситуации ошибки в Y распределены *равномерно* на некотором неизвестном симметричном интервале $[-S, S]$, то функция правдоподобия принимает вид $\prod_i \chi_S[V_i - \Phi(X_i, a)] / (2S)$, где $\chi_S(\cdot)$ – характеристическая функция указанного интервала. В этом случае МП-оценки параметров S и a даются решением оптимизационной задачи с минимаксным критерием

$$S = \max_i |V_i - \Phi(X_i, a)| \Rightarrow \min. \blacksquare \quad (2.6)$$

На практике распределение случайных ошибок наблюдений обычно считается нормальным. Это, однако, может иногда противоречить экономическому содержанию переменных. В этой связи представляет интерес рассмотреть ситуацию, когда распределение ошибок качественно отличается от нормального.

Пример 2.6. Макроэкономическая производственная функция имеет вид: $D = aK^\alpha L^{1-\alpha} e^{\gamma}$, где D – валовой внутренний продукт (ВВП), K – основной капитал, L – затраты труда (отработанные человеко-часы). Необходимо оценить параметры a , α и γ этой зависимости (γ часто интерпретируется как темп автономного технического прогресса). С учетом влияния прочих факторов на ВВП результаты наблюдений могут быть описаны регрессионной зависимостью $Y_i = aX_i^\alpha e^{\gamma} \xi_i$, где $Y = D/L$ – производительность труда, $X = K/L$ – фондовооруженность труда, а ξ_i – ошибка. Будем считать, что производственная функция выражает не фактический, а *максимально возможный* объем производства национального дохода при данных трудовых и капитальных ресурсах. Между тем влияние случайных факторов приводит к недоиспользованию ресурсов, т.е. к уменьшению ВВП против рассчитанного по модели. Это значит, что ошибки ξ_i меньше единицы и отражают степень использования "мощностей народного хозяйства". Примем поэтому, что ξ_i независимы и распределены на отрезке $[0, 1]$, причем функция их распределения имеет вид $P(\xi) = \xi^m$.

Тогда Y_i будут иметь на $[0, aX_i^\alpha e^{\gamma}]$ функцию распределения $\left(\frac{Y_i}{aX_i^\alpha e^{\gamma}} \right)^m$

и плотность $\frac{m}{Y_i} \left(\frac{Y_i}{aX_i^\alpha e^{\gamma}} \right)^{m-1}$. При этом критерий максимального правдоподобия примет вид:

$$\prod_i \frac{m}{Y_i} \left(\frac{Y_i}{aX_i^\alpha e^{\gamma}} \right)^{m-1} \Rightarrow \max, \quad (2.7)$$

однако величины, стоящие в квадратных скобках, при этом должны быть меньше единицы. Отсюда легко находится оптимальное значение a : $a = \max_i Y_i X_i^{-\alpha} e^{-\gamma}$. Теперь, логарифмируя (2.7), получаем следующее условие оптимальности для α и γ :

$$Q = \sum_s \left\{ \max_i [\ln Y_i - \alpha \ln X_i - \gamma] - [\ln Y_s - \alpha \ln X_s - \gamma] \right\} \Rightarrow \min.$$

Теперь из (2.7) можно найти и оптимальное значение m : $m = T/Q$.

В следующей таблице приведены данные по США за период 1946–1960 гг., рассчитанные по [4]:

Годы	1946	1947	1948	1949	1950	1951	1952	1953
$\ln(Y_i)$	0,372	0,350	0,392	0,419	0,489	0,504	0,522	0,534
$\ln(X_i)$	1,024	1,028	1,057	1,140	1,143	1,148	1,175	1,183

Годы	1954	1955	1956	1957	1958	1959	1960
$\ln(Y_i)$	0,563	0,593	0,596	0,610	0,598	0,637	0,654
$\ln(X_i)$	1,258	1,246	1,253	1,278	1,304	1,275	1,281

Расчет по приведенной модели дает: $\alpha = 0,568$, $\gamma = 0,00829$, $m = 50,2$. ■

Если имеется несколько теоретических моделей искомой зависимости, МП-метод позволяет не только оптимально оценить их параметры, но и выбрать лучшую из них (при этом важно, чтобы функции правдоподобия в разных моделях отражали плотности распределения одних и тех же переменных).

Пример 2.7. Имеются две модели зависимости между Y и X : $Y_i = aX_i + \xi_i$, где ξ_i – независимые нормально распределенные ошибки с нулевым средним (см. пример 2.4). В модели 1 дисперсия ξ_i постоянна, в модели 2 – пропорциональна X_i , соответственно будут различаться и функции правдоподобия и оптимальные значения a . При этом лучшей модели будет отвечать большее значение функции правдоподобия (по существу, это эквивалентно получению МП-оценки дискретного параметра i – номера модели; поэтому близким значениям параметра не обязательно должны отвечать близкие зависимости). Используя формулы примера 2.4, получим, что для этого надо сравнить

$$\left[\frac{2\pi e}{T} \sum_i (Y_i - aX_i)^2 \right]^{-T/2} \text{ и } \left[\frac{2\pi e}{T} \sum_i (Y_i - aX_i)^2 / X_i \right]^{-T/2} \left[\prod_i X_i \right]^{-1/2}.$$

При сравнении моделей 1 и 2 (с оптимальными для них параметрами a) было бы неверно сопоставлять сумму квадратов "обычных" отклонений

$\sum_i (Y_i - aX_i)^2$ в модели 1 с суммой квадратов взвешенных отклонений
 $\sum_i (Y_i - aX_i)^2 / X_i$ в модели 2. ■

МП-метод может использоваться и при неоднородности исследуемой совокупности.

Пример 2.8. Исследуется зависимость между производственной мощностью (P) предприятия и затратами (C) на его создание. Теоретический анализ [15] показывает, что эта зависимость степенная: $C = c_1 P^\alpha$, где α – подлежащий оценке фактор концентрации, отражающий процентный прирост затрат при увеличении мощности на 1%. Обозначив $\ln C = Y$, $\ln P = X$, $\ln c_1 = b$ и учтя влияние "прочих" факторов, представим эту зависимость как регрессионную: $Y = \alpha X + b + \xi$. Ошибки ξ будем предполагать независимыми и имеющими нормальное распределение с нулевым средним. Для определения α можно было бы применить метод наименьших квадратов (см. пример 2.5), используя данные Y_i и X_i для действующих предприятий. Однако можно получить и более точные результаты, если учесть, что влияние фактора концентрации должно отражаться и при проектировании новых предприятий. С учетом возможных просчетов в определении сметной стоимости объектов при их проектировании и влияния на эту стоимость факторов, не связанных с мощностью объектов, соответствующая модель должна быть записана в виде $V_k = \alpha U_k + d + \eta_k$, где V_k – логарифм сметной стоимости предприятия по k -му проекту ($k = 1, \dots, K$), U_k – логарифм его проектной мощности. Существенно, что значение α в обеих моделях одно и то же, хотя свободные члены и дисперсии S и R ошибок переменных ξ_i и η_k различны.

Критерий правдоподобия в данном случае относится ко всей совокупности фактической и проектной информации и имеет вид:

$$\prod_{i=1}^T \frac{1}{\sqrt{2\pi S}} \exp \left\{ -\frac{[Y_i - \alpha X_i - b]^2}{2S} \right\} \times$$

$$\prod_{k=1}^K \frac{1}{\sqrt{2\pi R}} \exp \left\{ -\frac{[V_k - \alpha U_k - d]^2}{2R} \right\} \Rightarrow \max.$$

Для оптимального α отсюда получается нелинейное уравнение, решаемое численными методами. Такой подход был применен в [15] к расчету влияния фактора концентрации производства на показатели предприятий сборного железобетона. ■

При использовании МП-метода возможен и *байесовский подход*, подробно рассмотренный в [1, 39]. Здесь предполагается, что оцениваемый параметр a является реализацией некоторого случайного вектора с известным априорным распределением.

Пример 2.9. Оценивается параметр a в зависимости $Y_i = aX_i + \xi_i$, где ошибки ξ_i независимы и имеют нормальное распределение с

нулевым средним и дисперсией bX_t , пропорциональной X_t . Предполагается, что априорное распределение величины a – показательное со средним значением m . Функция правдоподобия здесь имеет вид:

$$m^{-1} e^{-am} \prod_{t=1}^T \frac{1}{\sqrt{2\pi b X_t}} e^{-(Y_t - aX_t)^2 / 2bX_t}, \text{ и ее максимизация приводит к формам:}$$

$$a = \frac{V}{U} - \frac{mT}{2} + \sqrt{\left(\frac{mT}{2}\right)^2 + \left(\frac{V}{U}\right)^2 - \frac{W}{U}}; \quad b = \frac{1}{T} \sum_{t=1}^T \frac{(aX_t - Y_t)^2}{X_t},$$

где $U = \sum_{t=1}^T X_t$; $V = \sum_{t=1}^T Y_t$; $W = \sum_{t=1}^T \frac{Y_t^2}{X_t}$. ■

Пример 2.10. Рассмотрим ту же производственную функцию, что и в примере 2.6, представив ее в виде линейной зависимости $Y = \alpha X + \beta$, где Y – логарифм производительности труда, X – логарифм фондовооруженности труда. С учетом влияния прочих факторов результаты наблюдений могут быть описаны моделью $Y_t = \alpha X_t + \beta + \xi_t$, где ошибки ξ_t имеют нормальное распределение с неизвестной дисперсией S . Ставится задача оценки параметра α в этой зависимости. По экономическому содержанию $0 \leq \alpha \leq 1$, причем имеющиеся данные по другим странам позволяют считать априорное распределение этого параметра бета-распределением с плотностью $\frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} (1 - \alpha)^{\mu-1} \alpha^{\nu-1}$ и на-

более вероятным значением $\frac{\nu - 1}{(\mu + \nu - 2)}$. Тогда МП-критерий принимает вид:

$$\frac{\Gamma(\mu + \nu)}{\Gamma(\mu)\Gamma(\nu)} (1 - \alpha)^{\mu-1} \alpha^{\nu-1} \prod_{t=1}^T \frac{1}{\sqrt{2\pi S}} e^{-(Y_t - \alpha X_t - \beta)^2 / 2S} \Rightarrow \max. \quad (2.8)$$

Максимизируя по S , найдем: $S = \frac{1}{T} \sum_{t=1}^T (\alpha X_t - Y_t - \beta)^2$, после чего критерий (2.8) упрощается:

$$(1 - \alpha)^{\mu-1} \alpha^{\nu-1} \frac{1}{T} \left[\sum_{t=1}^T (Y_t - \alpha X_t - \beta)^2 \right]^{-T/2} \Rightarrow \max.$$

Отсюда для определения α получается уравнение 4-й степени, имеющее единственный корень $0 < \alpha < 1$. ■

2.6. Два правдоподобия

В классическом регрессионном анализе предполагается, что объясняемая переменная наблюдается с ошибкой, в то время как объясняющие переменные предполагаются измеренными точно. В этой ситуации функция правдоподобия в равной мере характеризует и правдоподобие

истинного значения переменной и правдоподобие ошибки наблюдения.

Между тем МП-метод используется и тогда, когда ошибки присущи наблюдениям нескольких переменных или когда количество случайных ошибок превышает количество наблюдений. При этом возникает двусмысленность, на которую в литературе, по-видимому, до сих пор не обращалось должного внимания. Дело в том, что в подобной ситуации неясно, правдоподобие *чего* должна отражать функция правдоподобия.

Пусть оценивается зависимость $Y = f(X, a)$ по известным результатам наблюдений $U_i = X_i + \xi_i$ и $V_i = Y_i + \eta_i$. В таком случае ошибки наблюдений связаны соотношением

$$\eta_i + f(U_i - \xi_i, a) - V_i = 0. \quad (2.9)$$

Здесь неясно, должна ли функция правдоподобия отражать *совместную* плотность распределения ошибок ξ_i и η_i (т.е. плотность вероятности того, что одна из ошибок будет равна ξ_i , а другая – η_i), либо она должна отражать плотность распределения *одной* случайной величины $\eta_i + f(U_i - \xi_i, a) - V_i$ (т.е. плотность вероятности выполнения равенства (2.9)). В первом случае МП-метод позволит нам установить наиболее правдоподобные значения ξ_i и η_i (и соответственно – истинных значений наблюдаемых характеристик), во втором же – только наиболее правдоподобное *множество* пар (ξ_i, η_i) удовлетворяющих ограничению (2.9). Таким образом, мы имеем два разных правдоподобия и соответственно два МП-метода. Условно можно назвать первый из них методом максимального правдоподобия ошибок (МПО-метод), а второй – методом максимального правдоподобия зависимостей (МПЗ-метод). Рассмотрим различия этих методов на примерах.

Пример 2.11. Случайная величина имеет равномерное распределение на отрезке $[0, b]$, т.е. имеет плотность распределения $b^{-1}\chi(x/b)$, где $\chi(x)$ – характеристическая функция отрезка $[0, 1]$. Для оценки неизвестного параметра b производятся T наблюдений. Результатом t -го наблюдения является сумма U_t двух независимых реализаций (ξ_t, η_t) указанной случайной величины. МПО-метод использует функцию правдоподобия совместного распределения ξ_t и η_t , что приводит к задаче

$$\prod_t b^{-2} \chi(\xi_t / b) \chi(\eta_t / b) \Rightarrow \max_t \quad (2.10)$$

при ограничениях $\xi_t + \eta_t = U_t$.

Легко видеть, что ее решением будет $b = \frac{1}{2} \max_t U_t$. Оптимальные ξ_t и η_t здесь не единственны, однако одной из оптимальных является комбинация $\xi_t = \eta_t = U_t / 2$.

При использовании МПЗ-метода необходимо вначале установить распределение результатов наблюдений $U_t = \xi_t + \eta_t$. Легко проверить, что их распределение треугольное с плотностью $p(x) =$

$= b^{-1} \max \{0; 1 - |b^{-1}x - 1|\}$. Поэтому оптимальное b здесь должно находиться из условия: $\prod_t b^{-1} \max \{0; 1 - |b^{-1}U_t - 1|\} \Rightarrow \max$. Нетрудно убедиться, что теперь оптимальное b всегда будет больше половины максимального из U_t . ■

Пример 2.12. Оцениваются параметры зависимости $Y = \Phi(X_1, X_2, \dots, X_n, a)$ по данным наблюдений Y_t и $U_{it} = X_{it} + \xi_{it}$, где ξ_{it} — независимые нормально распределенные ошибки с нулевыми средними и одной и той же неизвестной дисперсией S . Для применения МПЗ-метода необходимо определить плотность распределения случайной величины $\Phi(U_{1t} - \xi_{1t}, \dots, U_{nt} - \xi_{nt}, a)$. Эта задача сводится к вычислению n -кратных интегралов по областям, зависящим от вида функции Φ . Если Φ не линейна, это может представить значительную сложность.

В то же время при использовании МПО-метода подобных сложностей не возникает. Действительно, плотность совместного распределения наблюдений (функция правдоподобия) здесь равна

$\prod_{i,t} \frac{1}{\sqrt{2\pi S}} \exp\left\{-\frac{(U_{it} - X_{it})^2}{2S}\right\}$, поэтому решение задачи обладает следующими свойствами:

1) Для любого t точка (X_{1t}, \dots, X_{nt}) является ближайшей к (U_{1t}, \dots, U_{nt}) точкой изокванты $Y_t = \Phi(X_{1t}, X_{2t}, \dots, X_{nt}, a)$ в евклидовой метрике.

$$2) S = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (X'_{it} - X_{it})^2.$$

Оптимальная оценка a находится из условия $S \Rightarrow \min$, т.е. по предложенному в [6] из других соображений критерию минимального среднеквадратичного расстояния от наблюдаемых точек (U_{1t}, \dots, U_{nt}) до соответствующих изоквант функции Φ . ■

Пример 2.13. Оценивается параметр a зависимости $Y_t = X_t + a$, причем X_t и Y_t наблюдается с аддитивными ошибками, т.е. результатами наблюдений являются величины $U_t = X_t + \xi_t$ и $V_t = Y_t + \eta_t$. Теперь искомую зависимость можно представить так: $W_t = a - \xi_t + \eta_t$, где $W_t = V_t - U_t$. Примем, что ошибки ξ_t и η_t имеют плотности распределения $p(\xi, R)$ и $q(\eta, R)$, зависящие от неизвестного параметра масштаба R .

При МПО-методе параметры a и R вместе с ξ_t и η_t находятся путем решения задачи:

$$L_1 = \prod_t p(\xi_t, R) q(\eta_t, R) \Rightarrow \max; \quad \eta_t - \xi_t W_t - a \forall t.$$

При МПЗ-методе функция правдоподобия отражает плотность распределения $s(\zeta, R) = \int q(\eta, R) p(\eta - \zeta, R) d\eta$ разности $\zeta_t = \eta_t - \xi_t$, так что задача оптимизации параметров a и R принимает вид: $L_2 = \prod_t s(W_t - a, R) \Rightarrow \max$. ■

В примере 2.13 обе ошибки имели одинаковый "масштаб" (R). Между тем величина и "разброс" разных характеристик экономических объектов зависят от разных факторов. Поэтому хотелось бы использовать МП-метод и тогда, когда равенство (или пропорциональность) неизвестных дисперсий ошибок разных переменных не предполагается. Однако в этой ситуации МП-метод не всегда "работает".

Пример 2.14. Оценивается зависимость $Y_i = aX_i$, в условиях, когда вместо X_i и Y_i наблюдаются $U_i = X_i + \xi_i$ и $V_i = Y_i + \eta_i$, причем ошибки ξ_i и η_i независимы и имеют нормальные распределения с нулевыми средними и неизвестными дисперсиями Q и R . Легко видеть, что здесь $V_i = aU_i + a\xi_i - \eta_i$.

Решим эту задачу вначале МПЗ-методом. Для этого обозначим $\zeta_i = a\xi_i - \eta_i$ и заметим, что эта случайная величина имеет нормальное распределение с нулевым средним и дисперсией $a^2Q + R$. Теперь зависимость принимает вид $V_i = aU_i + \zeta_i$, и ей отвечает функция правдоподобия

$$L = \prod_i \frac{1}{\sqrt{2\pi(a^2Q + R)}} \exp \left\{ -\frac{(V_i - aU_i)^2}{2(a^2Q + R)} \right\}. \quad \text{Отсюда легко вытекает, что оптимальное } a \text{ определяется из условия}$$

$S = \sum_{i=1}^T (V_i - aU_i)^2 \Rightarrow \Rightarrow \min$, и при этом оптимальной оценкой для $a^2Q + R$ будет S/T . Оценить ξ_i и η_i и, тем более, Q и R в *отдельности*, к сожалению, здесь не удается.

Следуя [2, 40], применим теперь МПО-метод. Поскольку $\eta_i = aU_i + a\xi_i - V_i$, то здесь:

$$L = \prod_{i=1}^T \frac{1}{2\pi\sqrt{QR}} e^{-\frac{\xi_i^2}{2Q} - \frac{\eta_i^2}{2R}} = \prod_{i=1}^T \frac{1}{2\pi\sqrt{QR}} e^{-\frac{\xi_i^2}{2Q} - \frac{(aU_i - V_i + a\xi_i)^2}{2R}} \Rightarrow \max. \quad (2.11)$$

Приравнявая к нулю логарифмические производные L по ξ_i , a , Q и R , находим:

$$\frac{\xi_i}{Q} = -\frac{a(aU_i - V_i + a\xi_i)}{R}, \quad (2.12)$$

$$\sum_i (aU_i - V_i + a\xi_i)(U_i + \xi_i) = 0, \quad (2.13)$$

$$-\frac{T}{Q} + \frac{1}{Q^2} \sum_i \xi_i^2 = 0; \quad -\frac{T}{R} + \frac{1}{R^2} \sum_i (aU_i - V_i + a\xi_i)^2 = 0. \quad (2.14)$$

Как выяснится далее, следующее преобразование имеет решающую роль. Говоря словами [2. П. 29. 14], "уравнения (2.14) дают":

$$Q = \frac{1}{T} \sum_i \xi_i^2; \quad R = \frac{1}{T} \sum_i (aU_i - V_i + a\xi_i)^2. \quad (2.15)$$

Учтем теперь, что в силу (2.12) $\frac{1}{Q^2} \sum_i \xi_i^2 = \frac{a^2}{R^2} \sum_i (aU_i - V_i + a\xi_i)^2$.

Отсюда и из (2.15) получаем $\frac{T}{Q} = \frac{a^2 T}{R}$, так что $R = a^2 Q$ – это соотношение впервые получено в [40]. Однако соотношение истинных значений Q и R может быть иным, так что данный метод может давать *несостоятельные* оценки, что выяснено еще в [41]. Тем не менее для малых выборок несостоятельность оценок не слишком опасна, особенно, если учесть, что дисперсии ошибок – не самые важные параметры оцениваемых зависимостей.

Подставив $R = a^2 Q$ в (2.12), найдем: $\xi_i = \frac{V_i}{2a} - \frac{U_i}{2}$, после чего (2.13) дает: $\sum_i \left(\frac{U_i}{2} + \frac{V_i}{2a} \right) \left(\frac{aU_i}{2} - \frac{V_i}{2} \right) = 0$. Отсюда находим: $a^2 = \frac{\left\{ \sum_i V_i^2 \right\}}{\left\{ \sum_i U_i^2 \right\}}$. Легко проверить, что из двух значений a , удовлетворяющих этому условию, максимальное значение функции правдоподобия будет достигаться, если знак a совпадает со знаком $\sum_i U_i V_i$.

Полученная оценка a обладает двумя важными свойствами:

- 1) она является средним геометрическим из оценки $\sum_i U_i V_i / \sum_i U_i^2$, отвечающей уравнению регрессии $V_i = aU_i - \eta_i$, и оценки $\sum_i V_i^2 / \sum_i U_i V_i$, отвечающей "обратному" уравнению регрессии $U_i = a^{-1}V_i + \xi_i$;
- 2) она минимизирует критерий $|a|^{-1} \sum_{i=1}^T (V_i - aU_i)^2$.

Казалось бы, такое решение вполне удовлетворительно. Однако это не так! Дело в том, что проведенные преобразования некорректны, (2.14) отнюдь не всегда "дает" (2.15), а найденное сочетание ξ_i , a , Q и R не обеспечивает максимума функции правдоподобия (2.11). Действительно, положим все ξ_i равными нулю и устремим Q к нулю. Нетрудно убедиться, что в этом случае $L \rightarrow \infty$, тогда как при найденных выше ξ_i , a , Q и R величина L из (2.11) конечна. Таким образом, максимум функции правдоподобия в данной задаче *бесконечный* и достигается совсем при других сочетаниях ξ_i , Q и R . К тому же теперь становится неясным, какое значение a является оптимальным – любому a отвечает одно и то же бесконечное значение функции правдоподобия. Но, может быть найденное сочетание ξ_i , a , Q и R обеспечивает не глобальный, а локальный максимум функции правдоподобия? Оказывается, что и это не так: можно показать, что оно отвечает *седловой точке* функции правдоподобия – при отклонениях от нее в одних направлениях эта

функция убывает, но при отклонениях в других направлениях (например, при уменьшении Q и увеличении R) – растет. Некорректность же проведенных преобразований объясняется просто: равенства (2.15) являются следствиями (2.14) только если Q и R отличны от нуля! ■

Как видно из последнего примера, МПО-метод оказывается неработоспособным в некоторых ситуациях. При этом одного того факта, что число ошибок превышает число наблюдений, недостаточно: в рассмотренном примере этот метод привел бы к корректному решению, если бы было известно *отношение* дисперсий Q и R , но не позволил бы оценить a , если бы была известна *сумма* этих дисперсий.

2.7. Основные недостатки метода максимального правдоподобия

Мы видели, что МП-метод идейно прост, имеет широкую сферу применения и часто дает хорошие практические результаты. Однако он имеет и ряд недостатков.

1. Выбрать МП-методом лучшую из альтернативных моделей зависимости, как в примере 2.7, удастся не всегда. Рассмотрим две модели зависимости между характеристиками Y_t и X_t объекта в динамике. В соответствии с первой $Y_t = aX_t + b + \xi_t$, причем ξ_t – независимые нормально распределенные случайные величины с нулевым средним и неизвестной дисперсией. Во второй модели $Y_t - Y_{t-1} = a(X_t - X_{t-1}) + \eta_t$, где η_t устроены аналогично. Для каждой из моделей МП-метод применим, однако *сравнить значения правдоподобия* нельзя, ибо в первом случае величина Y_t случайная (и ей отвечает некоторая плотность распределения), тогда как во втором Y_t считается детерминированной (и отвечающая ей плотность *бесконечна*). Аналогичная ситуация часто имеет место при оценке параметров производственных функций.

2. На практике МП-метод обычно применяется для оценки параметрических зависимостей, когда функция правдоподобия зависит от *конечного* числа параметров. "В чистом виде" применить его в *непараметрическом* случае, по-видимому, нельзя (некоторые ситуации, где это удастся, рассматриваются в п. 4.2, однако они носят исключительный характер, и к тому же здесь не удастся оценить искомую зависимость во всей области ее определения). Далее, как уже отмечалось, иногда параметры оцениваемой зависимости можно рассматривать как реализацию некоторого случайного вектора с известным априорным распределением, применяя МП-метод в сочетании с байесовским оцениванием. Однако для непараметрических зависимостей такой подход реализовать не удастся, поскольку в пространствах функций нельзя ввести "хорошую" вероятностную меру, сосредоточенную на функциях, обладающих привычными для экономических моделей свойствами (например, достаточно гладких и выпуклых).

3. Пусть наблюдения величины Y имеют нормальное распределение с неизвестным средним a . Здесь наиболее правдоподобной и к тому же несмещенной оценкой параметра a является среднее арифметическое наблюдаемых значений показателя. Однако если функция распределения наблюдений даже незначительно отличается от функции нормального распределения (например, в равномерной метрике), среднее арифметическое перестает быть наиболее правдоподобной оценкой. Более того, эта оценка станет смещенной, и это смещение может быть сколь угодно большим. На практике это выражается в том, что наличие одного-двух резко выделяющихся наблюдений в большой выборке приводит к существенному искажению среднего значения. Трудности усиливаются в задаче построения регрессии – в [42] приведен пример, когда в выборке имеются два "сомнительных" наблюдения, причем вид линии регрессии существенно различен в зависимости от решения исследователя сохранить оба наблюдения, отбросить первое из них или отбросить второе.

Причину неустойчивости мы видим в том, что исходной информацией для МП-метода является вид плотности вероятностного распределения, в то время как этот метод слишком сильно реагирует на изменения этой плотности, особенно в той области, где она мала. Между тем даже тогда, когда совместное влияние большого числа малых случайных факторов обуславливает близкое к нормальному совокупное распределение отклонений, такая близость будет иметь место лишь вблизи центра распределения, тогда как на "хвостах" распределение далеко от нормального [43. Гл. XVI]. Однако поведение "хвостов" для МП-метода существенно: он придает чрезмерно большой "вес" большим ошибкам, информация о распределении которых наименее надежна. Эти и другие подобные трудности заставляют предъявить к методу оценки параметров вероятностных распределений дополнительное требование *устойчивости* (робастности, *robustness*), т.е. *малой чувствительности к малым отклонениям от исходных предположений* [42]. Сформулированное в таком общем виде, это требование естественно интерпретируется для любых типов задач оценки зависимостей и любых типов неопределенности.

По нашему мнению, сформулировав те или иные предположения о характере неопределенности наблюдений и выбрав подходящий метод оценки, каждый раз необходимо задаться вопросом: не может ли этот метод привести к серьезным ошибкам, если сделанные предположения будут "немного" нарушаться? Для получения на этом пути приемлемых методов оценки приходится вводить три типа ограничений:

- формализовать понятия "небольших отклонений" от исходных предположений (например, вводя расстояние между вероятностными распределениями и ограничивая величину этого расстояния) и "нечувствительности" к ним;

- ограничивать класс методов. Так, МП-оценкой параметра θ вероятностного распределения с плотностью $p(x, \theta)$ по наблюдениям

x_1, \dots, x_n будет решение уравнения $\sum_i p'_{\theta_i}(x_i, \theta) / p(x_i, \theta) = 0$. Однако более устойчивые оценки получаются [42], если ориентироваться на решение уравнений вида $\sum_i \psi(x_i, \theta) = 0$ с подходящими функциями ψ ;

- выбирать лучший метод в рассматриваемом классе, вводя дополнительные критерии качества оценки (например, минимум максимально возможного смещения при "наихудшем" отклонении закона распределения от нормального). При этом возникает проблема совмещения критериев точности и устойчивости оценок.

4. Мы уже отмечали принципиальное различие между дискретными и непрерывными распределениями: в первом случае функция правдоподобия выражает *вероятность* наблюдения тех или иных значений переменных, тогда как во втором – *плотность распределения* этих вероятностей. Рассмотрим подробнее, к чему может привести такое несоответствие в задаче оценки неизвестного параметра распределения.

Пусть случайная величина X принимает только целочисленные значения, а $p(n, \Theta)$ – вероятность события $X = n$. Имеются наблюдения n_1, \dots, n_T , о которых известно, что некоторые из них являются реализациями случайной величины X , тогда как остальные являются "грубыми ошибками" (резко выделяющимися наблюдениями). Известна, кроме того, вероятность ε появления грубой ошибки. Требуется оценить параметр Θ .

Допустим, что грубые ошибки являются реализациями некоторой случайной величины Y , при чем событие $Y = n$ имеет (неизвестную) вероятность $q(n)$. Обозначим через \mathcal{K} множество номеров грубых ошибок в выборке (n_1, \dots, n_T), а через k – количество таких ошибок. Вероятность того, что номера грубых ошибок образуют множество \mathcal{K} , а результатами всех ("правильных" и "грубо ошибочных") наблюдений будут соответствующие значения n_i , т.е. правдоподобие выборки, в этом случае составит $L = \varepsilon^k (1 - \varepsilon)^{T-k} \prod_{i \in \mathcal{K}} q(n_i) \prod_{i \notin \mathcal{K}} p(n_i, \Theta)$. Поскольку

$\sum_{i \in \mathcal{K}} q(n_i) \leq 1$, величина L будет максимальной при $q(n_i) = 1/k$ для всех

$i \in \mathcal{K}$. Тогда $L = (\varepsilon/k)^k (1 - \varepsilon)^{T-k} \prod_{i \in \mathcal{K}} p(n_i, \Theta)$. Никаких принципиальных

сложностей при максимизации этого выражения Θ и по \mathcal{K} здесь не возникает, и оптимальное решение всегда существует.

Ситуация в корне меняется, если распределения не дискретны. Здесь задача ставится аналогично: одни элементы выборки X_1, \dots, X_T являются реализациями случайной величины X с плотностью распределения $p(X, \Theta)$, другие являются грубыми ошибками, причем вероятность того, что результатом наблюдения окажется грубая ошибка, известна и равна ε . Однако теперь функцию правдоподобия невозможно даже записать, ибо распределение грубых ошибок может не иметь плотности (таковы ошибки из-за неправильного расположения

десятичной запятой в записи результата наблюдения) или иметь сколь угодно большую плотность в отдельных точках. Поэтому корректное решение задачи становится невозможным.

Чтобы получать разумные решения, приходится ставить задачу иначе. Так, в [42, 44] среди методов определенного класса ищется метод определения параметров искомой зависимости, дающий несмещенную оценку при отсутствии грубых ошибок и в некотором смысле наименьшее смещение оценки (математическое ожидание модуля или квадрата отклонения оценки от истинного значения) при наихудшем распределении грубых ошибок. Таким образом, и здесь не удастся уйти от проблемы совмещения разных критериев оценки.

2.8. Сочетание разных принципов

Практически приемлемые методы оценки непараметрических зависимостей можно построить, сочетая разные принципы оптимизации. Рассмотрим это на примере задачи интерполяции при наличии случайных ошибок в значениях функции – оценку функции $f(X)$ по результатам наблюдений $Y_i = f(X_i) + \xi_i$, где ξ_i – случайные ошибки, имеющие одно и то же нормальное распределение с нулевым средним. Здесь можно учесть одновременно требования малости интегрального отклонения и гладкости оцениваемой малости интегрального отклонения и гладкости оцениваемой функции. Приведем три примера возможных постановок:

а) отыскание "наиболее гладкой" зависимости при ограничении на степень правдоподобия случайных отклонений приводит к задаче:

$$Q(f) = \int |f''(X)|^2 dX \Rightarrow \min, \quad \sum_i (Y_i - f(X_i))^2 \leq D;$$

б) отыскание "наиболее гладкой" сепарабельной зависимости $f_1(X_1) + \dots + f_n(X_n)$ при аналогичном ограничении (ср. пример 2.1):

$$\sum_{i=1}^n \int_{-\infty}^{\infty} |f_i'(X_i)|^2 dX_i \Rightarrow \min;$$

$$\sum_i \left(Y_i - \sum_{i=1}^n f_i(X_{ii}) \right)^2 \leq D; \quad f_2(X_{21}) = f_3(X_{31}) = \dots = f_n(X_{n1}) = 0.$$

При таких постановках необходимо располагать информацией о среднем размере (например, о дисперсии) ошибок наблюдения;

в) использование критериального функционала, построенного инструментальным методом и учитывающего как "гладкость" зависимости, так и наблюдаемые отклонения от нее. Так, применительно к постановке а) можно было бы использовать критерий $\int |f''(X)|^2 dX + k \sum_i (Y_i - f(X_i))^2 \Rightarrow \min$, отразив в величине k "сравнительную важность" ошибок наблюдения и гладкости функции. .

Несмотря на определенный субъективизм, такие постановки приводят на практике к разумным решениям. Более того, оказывается возможным оценивать зависимости и при наличии случайных ошибок в наблюдениях объясняемой переменной. Так, если истинными значениями показателей t -го объекта ГС являются X_t и $Y_t = f(X_t)$, а наблюдаются величины $U_t = X_t + \xi_t$ и $V_t = Y_t + \eta_t$, содержащие случайные нормально распределенные ошибки с нулевым средним, можно оценить искомую зависимость, решая оптимизационную задачу $\int |f''(X)|^2 dX + \sum_i \{k_Y [V_i - f(X_i)]^2 + k_X [U_i - X_i]^2\} \Rightarrow \min$ с двумя "управляющими" параметрами k_X и k_Y . Особых вычислительных трудностей эта задача не представляет, и ее решением также оказывается некоторый сплайн.

2.9. Принцип максимальной согласованности

Рассмотренные выше принципы оценки зависимостей, несмотря на все их различия, допускают единую трактовку как частные случаи более общего **принципа максимальной согласованности** (далее – МС-принципа), в соответствии с которым оценка зависимости должна быть наилучшей с точки зрения критерия, отражающего степень ее "согласованности" (или "рассогласованности") с имеющейся информацией. Центральным при этом становится вопрос о разумном выборе измерителя "степени согласованности". Как показано выше, в этом качестве могут выступать и гладкость оцененной зависимости, и величина интегрального отклонения, и значение функции правдоподобия. Сам по себе МС-принцип выглядит достаточно тривиально. В том или ином смысле к возможно большей согласованности своих выводов с имеющейся информацией стремились почти все, занимающиеся поиском чего-либо: от исследователей до следователей. Поэтому данный принцип нельзя рассматривать как нечто абсолютно новое в науке. Однако, обобщая в нескольких направлениях принцип максимального правдоподобия, он позволяет расширить круг методов прикладной статистики. Например, становится очевидным, что структура критерия согласованности определяется характером исходной информации, а отнюдь не удобством вычислений или степенью разработанности математических методов.

При правильно выбранном критерии согласованности он, как и функция правдоподобия, может быть использован также для оценки той модели, которая была принята для описания функционирования объекта (см. пример 2.7). Поэтому выбор вида модели и выбор критерия оценки представляют собой части единого процесса и должны осуществляться согласованно [6]. В примере 2.7, где ошибки подразумевались случайными величинами, этот подход удалось реализовать в полном объеме. Однако возможна и иная ситуация. Пусть, например, влияние "прочих" факторов носит "интервальный" характер и имеются две

модели зависимости между Y и X . Первая модель постулирует, что отклонения $Y - aX$ по абсолютной величине не превосходят 10, вторая – что эти отклонения не превосходят $0,1Y$. Отыскание наиболее согласованных значений сводится здесь к решению системы неравенств $|Y_i - aX_i| < 10$ для первой модели и $|Y_i - aX_i| < 0,1Y_i$ – для второй. Прежде всего решение здесь может не существовать или не быть единственным. Отсутствие решения, скорее всего, свидетельствует о неадекватности модели. Однако если решения для обеих моделей существуют, но не единственны, все они оказываются "в равной степени" согласованными с имеющейся информацией и неясно, какая модель лучше. Причина этих трудностей – в специфической структуре критерия согласованности, который в данном примере является булевским (согласованность либо есть, либо ее нет). Поэтому надеяться на успешную практическую реализацию предложенного принципа можно лишь тогда, когда критерий согласованности достаточно информативен и чувствителен к изменениям оцениваемых параметров. В этой связи полезно остановиться на указанных в [1. Гл. 6] принципах "выбора общего вида функции регрессии".

С *первым* принципом – "максимальное использование априорной информации о содержательной сущности анализируемой зависимости" – нельзя не согласиться. Мы трактуем его как необходимость опираться на теоретическую модель рассматриваемого процесса или явления, поскольку только она и выражает указанную "априорную" информацию, а выражения типа "содержательная сущность зависимости" могут быть конкретизированы только в модельной форме. Более спорен *второй* принцип: выбор вида искомой зависимости должен производиться на базе "предварительного анализа геометрической структуры исходных данных", который, в соответствии с [1], состоит из анализа парных корреляционных полей и учета свойств гладкости искомой функции. Между тем в случае, когда число учитываемых факторов более двух, анализ парных корреляционных полей мало что дает, "зрительно" выявить геометрическую структуру исходных данных почти невозможно, а свойства гладкости искомой функции, по существу, формулируются в терминах характеристик теоретической модели, так что они должны быть учтены в соответствии с первым, а не вторым принципом.

Наконец, в качестве *третьего* принципа предлагается использовать статистические приемы, позволяющие "упростить" искомую зависимость. Особое внимание при этом обращено на преобразования, сводящие искомую зависимость к линейной. Трудно спорить, что *оценить* линейную зависимость обычно проще, чем нелинейную, хотя *выявить* линейный характер зависимости или догадаться о "виде ее нелинейности" удастся обычно только, когда в нее входит только одна объясняющая переменная. Однако в отрыве от теоретической модели "линеаризующие" преобразования обычно не имеют содержательного смысла. Если же теоретическая модель свидетельствует о наличии

зависимости, "сводимой" к линейной, это еще не означает, что соответствующее "сведение" упростит расчеты. Возьмем, например, зависимости $Y = \ln(aX + b) + \xi$, где ξ – нормально распределенная случайная величина с нулевым средним. Введя переменную $Z = e^Y$, мы придем к линейной зависимости Z от X , но случайная аддитивная добавка превратится при этом в мультипликативную, что не позволяет использовать методы "обычной" линейной регрессии.

Другая группа "статистических методов упрощения зависимости" состоит в отыскании "оптимального" количества учитываемых факторов. Однако если понимать под "факторами" какие-то характеристики объекта, то отбрасывание любого из них приводит к иной модели этого объекта, к оценке качественно иной зависимости. На каких-то стадиях исследования это, возможно, и допустимо, однако если исследователя интересует зависимость Y от X_1 и X_2 , то нельзя требовать, чтобы он изучал зависимость Y только от X_1 . Между тем, как видно из [1. С. 196–197], "факторами" предлагается считать не только X_1 , но и любые степени этой переменной. Если при этом коэффициенты при таких степенях имеют какой-то экономический смысл, то отбрасывание одного из факторов не позволяет установить важную характеристику объекта. Если же коэффициенты не имеют экономического содержания, то не имеет смысла и "окончательно полученная зависимость", независимо от величины интегрального отклонения, а процедура ее получения становится "игрой" с исходными данными.

Наконец, чрезвычайно важным и полезным представляется *четвертый* принцип: построенная зависимость должна быть возможно более устойчива по отношению к той совокупности исходных данных, на основании которой она оценена. Практическая реализация этого принципа предусматривает сопоставление оценок искомой зависимости, полученных на основании различных "подвыборок" из имеющихся наблюдений – "при неудачном выборе общего вида искомой зависимости результаты ее восстановления по различным выборкам, как правило, будут сильно отличаться один от другого". К сожалению, как-либо формализовать этот принцип затруднительно. Например, параметр α степенной зависимости $Y = kX^\alpha$ может быть надежно оценен, если в выборке есть наблюдения как с большими, так и с малыми значениями X . Если же оценивать параметры этой зависимости по подвыборке, включающей только большие X или только малые X , ошибки неизбежно будут велики. В других случаях надежная оценка параметра может потребовать, чтобы в выборке обязательно присутствовали наблюдения с большими, малыми и средними значениями X . Поэтому дать общие рекомендации по формированию "тестовых подвыборок" в общем случае трудно. Однако часто используют следующий метод: чтобы исключить ситуации, когда оценка искомой зависимости обусловлена каким-то одним наблюдением, из них отбрасывают сначала первое, потом второе и т.д. и определяют, сильно ли это повлияло на

окончательный результат (в [42. Гл. 7] приведен пример, когда вид зависимости принципиально меняется из-за отбрасывания одного наблюдения).

Критерий согласованности можно использовать не только для "точечной", но и для *интервальной* оценки параметров зависимостей. Здесь результатом оценки является некоторая доверительная область в пространстве параметров, в которой "с достаточно большой степенью реальности" будет лежать оцениваемый вектор a . Естественно включить в эту область все векторы a , для которых значение критерия согласованности составляет не менее такого-то процента от максимального. Такой подход является логичным обобщением известного принципа построения доверительных областей для параметров вероятностных распределений [45].

В критерий согласованности можно учесть и "потери". Действительно, пусть двум вариантам зависимости примера 2.3 ($Y = X$ или $Y = 4 - X$) отвечают соответственно разные конструктивные решения здания. Зависимости $Y = 4 - X$ отвечает более сложное и дорогое решение, и принятие такой зависимости, если на самом деле она неверна, приведет к перерасходу средств. Если принять зависимость $Y = X$, техническое решение будет проще и дешевле, однако если на самом деле имеет место зависимость $Y = 4 - X$, то построенное здание разрушится.

В подобной ситуации используются три способа рационального выбора:

- ограничивается (субъективно) снизу допустимый уровень правдоподобия получаемых зависимостей, а, следовательно, и множество "допустимых" сочетаний параметров оцениваемых зависимостей, после чего стараются минимизировать возможные потери от неправильного выбора параметров на этом множестве;

- выбираются "наиболее правдоподобные" сочетания неизвестных параметров, однако в последующие расчеты вводятся "коэффициенты запаса" (в расчетах конструкций – "коэффициенты запаса прочности"), устанавливаемые на основе прежнего опыта. Обратим внимание, что основная сфера применения этого подхода – технические расчеты, в экономических исследованиях он практически не используется;

- критерий оценивания формируется как агрегированный, учитывающий как правдоподобие оцененных зависимостей, так и потери, обусловленные их возможной неточностью. В [46] этот способ назван *инструментальным*.

Разумеется, широкое применение имеет и "волевой" способ, когда критерий оценивания не строится и не выбирается, а *принимается* без какой-либо мотивировки или со ссылкой на другие работы (где зачастую он также принимался без какого-либо обоснования). Применение такого способа, по нашему мнению, нецелесообразно, поскольку субъективизм в установлении критерия представляется априорно "большим злом", чем субъективизм в агрегировании (соизмерении) разнородных, но

естественных критериев. Заметим, что доказать ошибочность "волевого" способа, как правило, далеко не просто, хотя иногда это удается. Типичные примеры дает нам статистика. Так, традиционным в статистике стало исчисление среднего значения по формуле среднего арифметического, хотя хорошо известно, что при наличии грубых ошибок использование медианы дает большую точность [42, 44].

Как мы видели, методы оптимальных оценок, основанные на функциях правдоподобия, показателях интегрального отклонения, функционалах гладкости, имеют свою ограниченную сферу применения, тогда как хотелось бы иметь возможно меньшее число более универсальных критериев, с более широкой сферой применения. Далее будет показана возможность построения таких критериев, применимых и при нетрадиционных формализациях неопределенности результатов наблюдений.

2.10. Об учете целевой информации при оценивании зависимостей

Рассмотрим задачу установления параметров линейной зависимости $Y = aX + b$ между показателями Y и X промышленных предприятий некоторой отрасли. Наблюдаемые значения этих показателей для i -го предприятия в году t ($t = 1, \dots, T - 1$) образуют вектор (X_{it}, Y_{it}) . Предположим, что такая задача решается с целью "внедрения" ее результатов на конкретном крупном предприятии путем подбора рационального и приемлемого для него значения параметра X в следующем году T . Целевой характер такого исследования можно учесть, придавая в соответствующих расчетах больший вес последним годам отчетного периода и крупным предприятиям. Оправдан ли такой прием? Должна ли целевая информация учитываться дополнительно к исходной при построении критерия согласованности? Для ответа на эти вопросы заметим, что приведенная модель неполна: в ней отсутствует описание ошибок. Попробуем уточнить модель так, связь между наблюдаемыми показателями предприятия имеет вид $Y_{it} = aX_{it} + b + \xi_{it}$, где ξ_{it} – нормально распределенные случайные величины с нулевым средним. Однако и эта модель будет неполной, поскольку в ней не отражены закономерности, описывающие изменение ошибок "в пространстве и во времени". Между тем такие зависимости могут быть разными. Приведем ряд примеров.

1) Ошибки ξ_{it} в совокупности независимы и имеют одинаковую дисперсию. Легко видеть, что тогда оптимальные оценки параметров a и b находятся методом наименьших квадратов, причем всем наблюдениям приписывается один и тот же вес.

2) Ошибки ξ_{it} в совокупности независимы, но дисперсия ξ_{it} обратно пропорциональна мощности предприятия M_{it} в соответствующем году. Легко видеть, что здесь оптимальные оценки параметров a и b

находятся методом взвешенных наименьших квадратов, когда каждому наблюдению (X_{it} , Y_{it}) придается вес M_{it} . Такое решение, на первый взгляд, соответствует целевой информации. Предположим, однако, что модель случайных ошибок чуть-чуть иная: дисперсии ошибок *прямо* пропорциональны мощности предприятия. Легко видеть, что в такой ситуации веса наблюдений будут *обратно* пропорциональны мощностям, несмотря на наше желание придать больший вес крупным предприятиям. Этот пример, в частности, показывает, что при корректной формализации в модель могут войти и такие характеристики объектов ГС, которые непосредственно в искомой зависимости не фигурируют.

3) Все ошибки ξ_{it} имеют одинаковую дисперсию. Для разных предприятий ошибки независимы, однако для одного и того же предприятия ошибки разных лет скоррелированы. Примером может служить модель: $\xi_{it} = \beta \xi_{it-1} + \eta_{it}$, где η_{it} – независимые в совокупности нормально распределенные случайные величины с нулевым средним и одинаковой дисперсией. МС-принцип здесь требует выбрать a , b и β , максимизирующие соответствующую функцию правдоподобия. Однако это решение уже не обеспечивает минимума суммы квадратов отклонений с весами, зависящими от года, к которому относятся наблюдения.

В более общей ситуации для учета целевой информации вводятся "вспомогательные" характеристики объектов (M), отражающие "степень их соответствия целям исследования" и тем или иным способом "встраиваются" в критерий выбора оптимальных параметров искомой зависимости. Каким бы способом подобные операции не осуществлялись, за ними всегда скрывается явное или неявное представление исследователя о том, что объясняемая характеристика Y зависит не только от объясняющих (X), но и от "вспомогательных" переменных (M). Если такое представление явное, его можно формализовать в той или иной модели (типа приведенных выше). Хуже, когда исследователь вводит вспомогательные переменные и встраивает их в критерий оптимальности "на глазок", вместо того, чтобы включить их в состав объясняющих переменных модели и учесть выявленный предыдущими исследованиями характер их влияния на ту же объясняемую переменную Y .

Часто целевая информация может быть в достаточной мере учтена при формулировке и уточнении теоретической модели зависимости, т.е. на более ранних стадиях исследования, когда первоначальное представление об искомой зависимости и о неопределенности входящих в нее характеристик еще только формализуется.

3. ХАРАКТЕРИЗАЦИЯ НЕОПРЕДЕЛЕННОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

Гораздо легче найти ошибку, нежели истину.

И.В. Гёте

В п. 1.2 была представлена попытка классификации типов ошибок наблюдения в зависимости от их происхождения. В этом разделе основное внимание уделено различным подходам к формализованному описанию ошибок, их представлению в виде различных математических объектов.

3.1. Систематические ошибки

Наиболее простая причина несоответствия истинного и наблюдаемого значений характеристики – "систематические" ошибки, одинаковые при наблюдении всех значений какого-либо показателя. Они часто возникают в технике вследствие неправильной градуировки измерительного прибора. Рассмотрим случай оценки параметрической зависимости $Y = aX + b$ по данным наблюдений. При точных наблюдениях найти искомые коэффициенты можно всего по двум точкам.

Пусть, однако, все наблюдения Y содержат одну и ту же (постоянную) аддитивную ошибку. Легко видеть, что при этом коэффициент b не удастся точно определить ни при каком числе наблюдений. Однако, если бы речь шла о зависимости $Y = aX^2 + bX$, то хватило бы трех наблюдений, чтобы определить неизвестные a и b .

Таким образом, наличие систематических ошибок ограничивает класс зависимостей, которых можно оценить достаточно точно при достаточно большом объеме исходной информации. В общем случае наличие систематических ошибок лишь увеличивает количество параметров в задаче. Так, если искомая функция имеет вид $g(X, a_1, a_2, \dots)$, где g – известная функция, a_1, a_2, \dots – неизвестные параметры, а переменная X измеряется с систематической аддитивной ошибкой d , то эта задача эквивалентна отысканию функции $g(X + d, a_1, a_2, \dots)$, при условии, что переменная X измеряется без ошибок.

Для ситуации, когда результатом наблюдения является не число, а некий иной объект, можно предложить следующее определение, под которое в одномерном случае подпадают и аддитивные и мультипликативные систематические ошибки. Пусть V – пространство характеристик, которому принадлежат и результаты наблюдений, $F: V \rightarrow V$ –

некоторое взаимно-однозначное отображение. Будем говорить о наличии систематической ошибки наблюдений, если истинная характеристика объекта Y и результат ее наблюдения N связаны соотношением $N = FY$ (каждый раз вместо истинного значения характеристики наблюдения ее образ при отображении F).

3.2. "Случайные" результаты наблюдений – вероятностный подход

Для характеристики ошибок наблюдений обычно используется термин "случайные ошибки" или "ошибки, вызванные случайными факторами". Величина их колеблется, они могут быть положительными и отрицательными, и ошибки в одних переменных никак не связаны с ошибками в других (по крайней мере это может быть не видно "невооруженным глазом"). В дальнейшем иногда потребуются различать эту "практическую", разговорную трактовку понятия случайности от "математической", выражаемой в терминах теории вероятностей. Поэтому там, где это существенно, мы будем в первом случае брать слово "случайный" в кавычки, а во втором – использовать термины "стохастика" и "стохастический".

Одним из источников "случайных" ошибок может быть *индикация* – замена точной методики измерения показателя другой, приближенной (например, в силу ограниченности времени или ресурсов). Другой источник, по существу рассмотренный в п. 1.5.2, – наличие "прочих", не отраженных в модели факторов, влияние которых приводит к колебаниям наблюдаемых значений характеристик от истинных. Большинство статистиков не только называет эти колебания "случайными", но и рассматривает их как стохастические. Тем самым изменения методики измерения или влияние "прочих" (несущественных для исследователя) факторов рассматриваются как реализации каких-то случайных событий и объекты применения теоретико-вероятностных методов. Насколько это корректно?

Большинство событий, которые конкретный человек рассматривает как "случайные" (речь идет о событиях, являющихся предметом традиционного экономического анализа, а не о явлениях на, скажем, молекулярном уровне), предстают как закономерные, если взглянуть на них с иной точки зрения, учтя дополнительные характеристики и обстоятельства явления. Так, исход бросания конкретной монеты может быть рассчитан, исходя из скорости, с которой была брошена монета, силы ветра и направления броска. Опоздание на работу из-за задержки городского транспорта становится совершенно закономерным, если знать, что физический износ такого-то узла такого-то транспортного средства обусловил его поломку как раз в данное время. Такое представление о случайных событиях давалось еще в [47. С. 85–86]: "События, возникшие при встрече или комбинации явлений, принадлежащих независимым

рядам, получившимся в результате причинности, мы называем явлениями случайными или результатами случая". В более современных терминах это сформулировано в [48. С. 49]: "Случайное событие возникает тогда, когда в закономерное течение явления, определяемого данной причиной, вклинивается действие, определяемое другой причиной, не связанной с первой, и происходит воздействие одного явления на другое".

Примерно так же трактуется случайность результатов наблюдений в общей теории статистики: "При определенных условиях величину отдельного элемента совокупности можно рассматривать как случайную величину, имея в виду, что она является не только автоматическим результатом какой-то общей закономерности, но в то же время и сама определена действием множества факторов, не зависящих от этой общей закономерности" [37. С. 13].

Эти определения не являются строгими, но какое-то представление о генезисе случайности в природе, техносфере и обществе они, безусловно, дают.

Результаты наблюдений являются неопределенными именно потому, что в момент наблюдения нельзя точно описать ни условий наблюдения, ни конкретных особенностей методики наблюдения, которые приводят к отклонениям наблюдаемого значения характеристики от "истинного".

При вероятностном подходе эти факторы трактуются как условия некоторого "опыта". Предполагается, что такой опыт допускает многократные независимые повторения и при каждом повторении мы можем получить различные "элементарные исходы" (и соответственно различные результаты наблюдения), характеризующиеся определенной вероятностью.

Для обсуждения правомерности применения вероятностного подхода и последующего рассмотрения возможных альтернатив ограничимся дискретным случаем и приведем некоторые базовые положения теории вероятностей, относящиеся к случайным (в нашей терминологии – стохастическим) величинам (СВ).

П1. СВ X характеризуется некоторым счетным множеством (скалярных или векторных) неравных между собой значений r_i (точек в пространстве "элементарных событий"), а также набором величин p_i , трактуемых как "вероятности" или "степени реальности" событий¹⁴ $X = r_i$.

П2. Величины p_i неотрицательны и не превосходят единицы.

П3. СВ X отождествляется с детерминированной величиной r , если событию $X = r$ приписана вероятность 1, а событиям $X = s$ при $r \neq s$ – нулевые вероятности.

П4. Функция $Y = f(X)$ от СВ определяется следующим образом: это

¹⁴ В общем случае под "событием" понимается множество (объединение) элементарных событий.

другая СВ с возможными значениями $f(r_i)$. Поскольку при разных i числа $f(r_i)$ могут быть равными, то в общем случае вероятность события $Y = s$ определяется как сумма вероятностей событий $X = r_i$ по всем таким i , для которых $f(r_i) = s$.

Пусть \mathcal{A} – некоторое множество элементарных событий. Рассмотрим функцию $f(r)$, равную 1 при $r \in \mathcal{A}$ и 0 в противном случае. Тогда из определения П4 получим, что вероятность события $X \in \mathcal{A}$ равна сумме вероятностей элементарных событий, образующих множество \mathcal{A} (так, вероятность того, что величина X окажется равной r_1 или r_2 , будет равна $p_1 + p_2$). В частности, если множество \mathcal{A} включает все элементарные события, отсюда следует, что $\sum_i p_i = 1$.

П5. СВ X_1, \dots, X_k называются *независимыми*, если векторная СВ $Y = (X_1, \dots, X_k)$ такова, что при любых r_1, \dots, r_k вероятность события $Y = (r_1, \dots, r_k)$ равна произведению вероятностей событий $X_1 = r_1, \dots, X_k = r_k$.

На основе этих определений оказывается возможным ввести определения других действий над СВ. Например, результат операции суммирования независимых СВ $X + Y$ вводится как функция от компонент векторной СВ (X, Y) с независимыми компонентами.

Такая трактовка "случайности" позволяет без внутренних противоречий и порочных кругов формализовать представления и о равновероятности событий и, особенно, об их независимости (что Курно трактовал как "принадлежность явлений к независимым рядам", а Карпенко – как "несвязанность причин" различных явлений). Одновременно появляется возможность использовать модели теории вероятностей для обработки информации о случайных событиях или явлениях. Но какое отношение имеют полученные на этом пути результаты к реальной действительности?

Обычно в обоснование правомерности (а то и необходимости) вероятностной трактовки реальных процессов и явлений приводятся следующие доводы.

1. Многие экономические события или явления можно рассматривать как повторяющиеся, массовые, являющиеся предметом теории вероятностей. При статистических исследованиях часто выявляется наблюдаемая устойчивость их частот, что как раз и доказывается в теории вероятностей, равно как и близость к нормальному распределению суммы большого числа независимых малых СВ. Наблюдаемые значения показателей также можно рассматривать как повторяющиеся (*мысленно* можно представить повторное наблюдение с иными ошибками измерения или при иных значениях прочих факторов), а следовательно – моделировать как реализации какой-то СВ. Исследуемую группу объектов или наблюдаемую ее часть на этом основании можно рассматривать как случайную выборку из некоторой, также *мысленно* представляемой ГС, включающей как реально существующие, так и "потенциально возможные" объекты. Например, количество предприя-

тий данной отрасли конечно, но наблюдаемые значения показателей некоторых (или даже всех) из них рассматриваются как выборка из бесконечного числа показателей, которыми могли бы характеризоваться такие предприятия при данных условиях, если бы "прочие факторы" приняли другие возможные значения. Соответственно появляется возможность оценивать точность статистических процедур, имея в виду колебания оценок при повторных выборках из той же ГС или при повторных наблюдениях тех же объектов.

2. В теории полезности [49, 50] доказано, что при определенных предположениях неопределенность исходной информации может быть учтена путем введения некоторых параметров, интерпретируемых как "субъективные" вероятности на множестве возможных значений исследуемых показателей, что автоматически возвращает нас "в лоно" вероятностных моделей.

3. Во многих случаях вероятностная трактовка "случайных" ошибок ведет к разумным с практической точки зрения решениям и статистическим выводам. Так, подтвердили свою эффективность на практике методы статистического контроля качества продукции, основанные на теоретико-вероятностных предположениях.

4. Методы математической статистики позволяют (в вероятностных терминах) оценить ошибку полученных в результате расчетов значений параметров искомых зависимостей. При этом для "хороших" (состоятельных) методов ошибка определения нужного параметра (в параметрической зависимости между Y и X) стремится к нулю при неограниченном увеличении числа наблюдений.

5. Использование вероятностного подхода и аппарата является своего рода традицией, от которой многим трудно отказаться, в то время как альтернативные подходы требуют не только апелляции к новому математическому аппарату, но и известной научной смелости.

Рассмотрим, насколько убедительны эти доводы.

1. Прежде всего вызывает возражения объяснение понятия ГС как совокупности не только реальных, но и "потенциальных", воображаемых объектов. Возможность даже мысленного пространственного или временного повторения не всегда может быть содержательно оправдана. Опираясь на "повторяемость" (воспроизводимость) исследуемых явлений, прикладная статистика лишается возможности анализа уникальных (неповторяющихся в принципе) событий и явлений, выявления присущих им закономерностей. К тому же вероятностный подход предполагает независимость воспроизводимых наблюдений. Между тем если при разных наблюдениях используются одни и те же измерительные приборы, эксперты, формы отчетности и указания по их выполнению и т.п., то допускаемые при этом ошибки вряд ли можно априорно считать независимыми. Иными словами, математические представления о независимости могут расходиться с "обычными".

2. Действительно, субъективные вероятностные меры позволяют получать непротиворечивые и "разумные" решения. Однако этот метод не гарантирует воспроизводимости получаемых результатов (другой

исследователь, решая ту же задачу, может ввести другую субъективную меру). "Объективизировать" эти вероятности можно было бы в рамках некоторой процедуры стандартизации, аналогичной тому, как устанавливаются обязательные для проектировщиков коэффициенты запаса прочности конструкций. "Мир субъективных вероятностей" по необходимости субъективен и уникален, и согласование этих миров для различных соприкасающихся ситуаций и разных субъектов более чем проблематично.

3. Вероятностный подход действительно дает разумные практические результаты, но только там, где неопределенность носит "технологический" или "природный" характер. Его применимость к процессам и явлениям, где влияние "человеческого фактора" существенно, в общем случае проблематична. Так, в "азиатском" финансовом кризисе 1997 г. инвесторы, использующие вероятностные модели фондового рынка, пострадали не меньше, чем те, которые ими не пользовались.

4. Методы математической статистики действительно позволяют оценивать точность определения параметров зависимостей при наличии стохастических ошибок. Это, однако, не является отличительной особенностью вероятностного подхода, ибо методы оценки точности решения задач имеются и в других разделах математики (например, в теории устойчивости дифференциальных уравнений), а отсутствие таких методов не лишает, например, языкознание статуса научной дисциплины.

К тому же эффективность математико-статистических методов обосновывается только "в среднем". Однако, применяя хороший "в среднем" метод при решении конкретной статистической задачи, мы не можем быть уверены, что наша ситуация "средняя" (грубо говоря, если метод дает небольшую ошибку в 90% случаев, мы не можем быть уверены, что не попали в "плохие" 10%). Поэтому может оказаться, что найденные этим методом зависимости справедливы для указанной "генеральной совокупности", но не для той ее части, которая состоит из "реальных объектов". Не следует преувеличивать и важность того, что при увеличении объема выборки ошибка оценки искомой зависимости в определенном смысле ("с вероятностью 1") стремится к нулю. Действительно, не видим же мы ничего страшного в том, что большое количество малых воздействий может существенно изменить траекторию движения материальной точки, а большое количество опечаток может существенно изменить не только впечатление о книге, но и ее содержание. Поэтому отнюдь не всегда увеличение объема исходной информации должно компенсировать ее неточность.

5. Что касается традиции, то эта сторона дела имеет существенное значение, когда речь идет о давно сложившейся отрасли знания, обладающей свойствами преемственности полученных результатов (математика, физика, история и т.д.). К сожалению, экономическая наука в целом и экономико-математическое моделирование в частности не могут пока быть отнесены к такого рода дисциплинам. Многие базис-

ные положения экономической теории подвергаются регулярному пересмотру, а рекомендации приводят к провалам [51]. Та же неустойчивость характерна и для модельной части экономической науки, достаточно вспомнить в качестве примера многолетнюю дискуссию о применимости модели САРМ для описания функционирования фондового рынка [52].

Таким образом, наиболее распространенные доводы в пользу применения исключительно вероятностного подхода не следует считать неопровержимыми. Целесообразно поэтому более внимательно рассмотреть основы этого подхода и связь их с реальностью, имея в виду отнюдь не "ниспровержение" вероятностных методов (такие попытки хорошо известны), а более точное для сегодняшних условий определение места вероятностного подхода в арсенале современных средств моделирования экономики.

Начнем с того, что теория вероятностей как математическая наука базируется на определенной системе аксиом, при этом понятия "случайного события" и "вероятности" считаются первичными, неопределяемыми¹⁵. Таким образом, эта теория не выводит указанные понятия из свойств реального мира. Приведенные же выше и иные описательные определения "случайности" как свойства реального мира поясняют ее в иных терминах, которые сами нуждаются в определениях ("встреча или комбинация явлений", "другая причина, не связанная с первой" и т.п.). В общем случае одни понятия определяются на основе других, поэтому в конечном счете приходится принимать всю совокупность понятий без точного определения. Более того, базисное исходное понятие должно характеризовать некоторое свойство случайных событий, присущее реальному миру, и тем самым представлять некоторое утверждение, принимаемое без доказательства, "аксиому А", из которой все остальные утверждения выводятся дедуктивными рассуждениями. Но тогда, как показано в [53. С. 59–60], "поскольку абсолютная достоверность событий невозможна, все, что А могла бы утверждать, должно быть сформулировано только в терминах "вероятности".

Иными словами, теоретически определить вероятности, не ссылаясь на них, нельзя. Поэтому на практике исходная вероятностная мера (величины p_i), вводимая аксиоматически, возникает "из ниоткуда", и потому не отражает каких-то реально существующих объектов или явлений. Более того, она *неизмерима* (в нашей терминологии), а правомерность ее использования подтверждается (и то – с некоторой вероятностью!) лишь в условиях массового применения получаемых результатов.

Недостаточная наглядность понятия вероятностной меры, отсутствие его "материальных носителей" еще ярче проявляются при рассмотрении недискретных СВ. Так, трудно "наглядно" объяснить, почему из

¹⁵ Исходным в теории вероятностей считается в общем случае *вероятностное пространство*, включающее одновременно пространство элементарных событий, σ -алгебру подмножеств этого пространства и вероятностную меру, определенную на таких подмножествах.

того, что событие имеет вероятность 0, не следует его неосуществимость, а из того, что событие имеет вероятность 1, не следует, что оно обязательно осуществится¹⁶. Нельзя "наглядно" объяснить и тот факт, что не всякое множество элементарных исходов имеет вероятность. Разумеется, все эти факты становятся естественными и объяснимыми, если рассматривать их как логические следствия из первоначальных аксиом теории вероятностей. Но тогда мы имеем дело с ветвью чистой математики, развивающейся только на основе своих внутренних законов и оперирующей с удобными для их описания понятиями, не имеющими прямых аналогов в реальной действительности. Об этом весьма строго высказывается Дж. Литтлвуд [53. С. 62–63]: «Мы подходим, наконец, к связи между идеальной теорией и реальным миром, или "реальной" вероятностью. Сторонник математической школы, если он последователен, в вопросах приложений умывает руки. Тому, кто ими интересуется, он говорит, что реальная система существует параллельно обычной теории: "Если хотите, попробуйте: в мою задачу не входит обоснование приложений; это вопрос философский, я же математик". Он часто также слонен сказать: "попробуйте, если что-нибудь выйдет, то это и будет оправданием". Но теперь он уже не только философствует, но и делает характерную ошибку. Опытная проверка того, что гипотеза "работает", не может как индуктивное суждение служить обоснованием этой гипотезы по существу».

Таким образом, трактуя некоторый показатель экономического объекта как СВ, исследователь не может достаточно обоснованно ссылаться на какое-то свойство реальности, отражаемое вероятностной мерой. Однако он может поступить точно в соответствии с теоретико-вероятностным определением СВ, конкретно указав множество возможных значений показателя и вероятностную меру на нем. Но тогда вероятностная мера становится выражением для *представления исследователя о неопределенности рассматриваемого показателя* – из объективной вероятности она сейчас же превращается в субъективную. Так, исследователь может приписать двум сторонам монеты вероятности $1/2$ и назвать ее правильной (или сделать это со ссылкой на других авторов), однако никакие наблюдения за монетой не будут обоснованиями такого решения. Более того, он может проверить гипотезу, что монета правильная, но полученное решение также выражается в терминах вероятностей – получить достоверный ответ о правильности реальной монеты на этом пути невозможно.

Поэтому, строго говоря, вероятностная трактовка "случайности" основывается не на каких-то свойствах, присущих рассматриваемому реальному объекту, а на собственных (или почерпнутых у других авторов) представлениях исследователя о возможности описания поведения этого объекта вероятностными моделями. Однако, несмотря на невоз-

¹⁶ Например, что при реализации случайной величины ξ , равномерно распределенной на отрезке $[0, 1]$, было получено некоторое число b . Событие $\xi = b$ осуществилось, однако в соответствии с теорией вероятностей его вероятность равна нулю.

возможность строго дедуктивного доказательства стохастичности событий и процессов¹⁷, использование соответствующих гипотез в ряде случаев оказывается полезным и приводит к разумным практическим рекомендациям и результатам.

Рассмотрим, как выглядит теоретико-вероятностный подход к задаче оценки регрессионной зависимости $Y = f(X)$ между показателями некоторого экономического объекта в условиях, когда на самом деле Y зависит не только от компонент вектора X , но и от других характеристик объекта ("прочих" факторов). Поскольку при теоретическом анализе ни один из "прочих" факторов не был включен в модель, это означает, что влияние каждого из них на Y теория оценивает как слабое. Поэтому отклонение $Y - f(X)$ обусловлено совокупным влиянием большого числа факторов, каждый из которых слабо влияет на Y . На следующем шаге делаются три допущения:

1) влияние каждого из прочих факторов сводится к аддитивной добавке к $f(X)$, обращающейся в нуль при "средних" значениях прочих факторов;

2) указанные добавки являются СВ;

3) СВ, отвечающие разным "прочим" факторам, независимы или слабо зависимы.

Теперь в силу центральной предельной теоремы [43] отклонение $Y - f(X)$ может рассматриваться как СВ, имеющая близкое к нормальному распределение с нулевым средним, и задача сводится к одновременному нахождению (по данным наблюдений) искомой функции $f(X)$ и дисперсии этого распределения.

Наиболее дискуссионным в этой конструкции является второе допущение, которое, как отмечалось, в принципе не может быть достоверно подтверждено. Причины методологических затруднений, связанных с трактовкой отклонений $Y - f(X)$ как СВ, вряд ли лежат в конкретной системе аксиом, на которой базируется теория вероятностей. Во всяком случае, предпринятые до сих пор попытки не смогли преодолеть отмеченного выше разрыва между математическим понятием вероятности и свойствами реальных объектов. Подчеркнем еще раз, что такие соображения, как "устойчивость частот", возможность неограниченного повторения опытов, независимость отдельных наблюдений не могут являться основаниями для применения вероятностного подхода — наоборот, только после "аксиоматического" принятия этого подхода они могут выступать (но могут и не выступать) в качестве дополнительной информации о наблюдаемых значениях показателей.

В этой связи сфера применения вероятностного подхода представляется нам ограниченной. Он считается применимым при анализе таких явлений, как массовое обслуживание и отказы оборудования, однако уже при анализе колебаний биржевых котировок трудно отделить стохастические колебания от целенаправленной биржевой игры аген-

¹⁷ Например, не может быть достоверно доказано, что результаты выпадения монеты являются СВ в том понимании, какое дается в теории вероятностей.

тов, располагающих большими средствами. Точно так же трудно убедить кого-либо из сведущих специалистов в стохастике курса доллара на ММВБ в марте–июле 1995 г. Попробуйте, например, с позиций "несведущего" экономиста оценить ("интерполировать") этот курс на 15.04 1995 г. (фактически он равнялся 5046 руб.) по данным на 01.03 1995 г. (4499 руб.) и 01.08 1995 г. (4405 руб.).

Отметим также, что некоторые статистики отвергают вероятностный подход, видя задачу статистики лишь в анализе совокупности наблюдаемых данных, получении ее обобщенных, "средних" показателей [18, 19]. Модельному описанию ошибок наблюдений при этом не уделяется должного внимания (они рассматриваются только как "шум", мешающий точно оценить нужные параметры, но не как проявление общей для всей ГС закономерности). Поэтому критерий согласованности получаемых результатов с данными наблюдений приходится устанавливать каждый раз заново, руководствуясь эвристическими соображениями, а получаемые выводы оказываются достаточно "бедными". К тому же при таком подходе статистические расчеты требуют большого числа разнообразных исходных предположений, что сдвигает разумный баланс между исходной и выходной информацией в сторону превалирования входной информации. Построение статистической модели при этом зачастую оказывается "экономически" неэффективным. Такой "сверхобъективный" подход можно считать полярным по отношению к вероятностному.

Однако есть и промежуточные подходы, образующие совместно с приведенными выше относительно богатый спектр возможностей при выборе инструментария моделирования зависимостей. Для этого, по существу, необходимо предложить альтернативные трактовки величин отклонений $Y - f(X)$. В этих целях рассмотрим еще раз основные свойства П1–П5, характеризующие СВ, и попробуем выяснить, можно ли их заменить какими-то иными.

Каждая СВ X (в дискретном случае) задается своим множеством возможных значений (*носителем*), в данном случае состоящим из счетного числа точек r_i , и набором чисел p_i , характеризующих "степень реальности" каждого из этих значений. Такая характеристика представляется естественной и разумной, позволяя отличать одну СВ от другой по "степени реальности" их значений. Естественно и удобно также считать, что "степень реальности" лежит между 0 и 1. Вполне разумно и определение детерминированных величин, позволяющее рассматривать их как частный случай СВ. Таким образом, положения П1–П3 каких-либо серьезных возражений не вызывают. Однако положения П4–П5 отнюдь не очевидны и имеют достаточно разумные альтернативы, так что ориентация эконометрики исключительно на теорию вероятностей может рассматриваться скорее как дань традиции, чем объективная необходимость.

3.3. Альтернативные описания "случайных" величин

Как отмечалось, для того, чтобы учесть неопределенность результата наблюдения, ее характер, необходимо уточнить и формализовать, описать с помощью некоторого математического объекта. При вероятностной трактовке таким объектом является "вероятностная мера" или "вероятностное распределение". Это означает, что при известном истинном значении показателя X результат его наблюдения N рассматривается как СВ с заданным (однозначно или с точностью до некоторых неизвестных параметров) вероятностным распределением $p(X, N)$, характеризующим "степень реальности" реализации наблюдения N (в дискретном случае это вероятность того, что наблюдаемое значение показателя X окажется равным N) и удовлетворяющим требованиям, накладываемым теорией вероятности на вероятностные распределения. Однако такая формализация неопределенности не является единственно возможной. Возможны и иные формализации неопределенности, когда результат N наблюдения показателя X характеризуется "степенью реальности" $p(X, N)$, лежащей между 0 и 1, однако понятия независимости неопределенных величин и операции над ними вводятся иначе. Таким трактовкам "случайности" будут отвечать иначе методы оценки зависимостей.

В литературе известны два варианта замены приведенных выше аксиом (положений) П4 и П5. Первый вариант – так называемые **нечеткие** величины (НВ). Этому вопросу (теории нечетких множеств, нечетких подмножеств или размытой логике) посвящена большая литература, например, [54–59]. Так же, как и СВ, НВ X может принимать различные векторные (в частном случае – скалярные) значения r , и событие $X = r$ так же характеризуется "степенью реальности" его осуществления $p_X(r)$, лежащей в пределах от 0 до 1. В литературе число $p_X(r)$ именуется "степенью принадлежности r к множеству возможных значений X ". Далее мы будем иметь дело только с такими НВ, для которых $\sup_r p_X(r) = 1$ – в [55, 56] они именуется *нормальными*.

Функция $p_X(r)$, определенная на всем множестве возможных значений НВ X , называется "функцией принадлежности". Как и СВ, НВ X трактуется как детерминированная величина r , если $p_X(r) = 1$, $p_X(s) = 0$ при всех $s \neq r$. Функции принадлежности являются такими же "первичными" характеристиками НВ, как функции распределения – для СВ. В конкретных приложениях, как и в случае теории вероятностей, эти функции задаются экспертно, либо преобразуются по определенным правилам при выполнении тех или иных операций над НВ (при любых способах учета неопределенности уйти от ее формализации не удастся). Главные отличия НВ от СВ заключаются в ином способе определения операций с НВ и их независимости.

Именно, функция $Y = f(X)$ от НВ (так же как и для СВ) опреде-

ляется следующим образом: это другая НВ, возможными значениями которой являются $f(r_i)$. При этом уравнение $f(r) = s$ может вообще не иметь решений, иметь одно или много решений r . Принимается, что степень $p_Y(r)$, принадлежности вектора s к множеству значений НВ Y в первом случае равна 0, во втором – $p_X(r)$, в третьем – *наибольшей* (точнее – *верхней грани*) из соответствующих величин $p(r)$. Разумеется, как и в вероятностном случае, для НВ также можно определить понятие события как некоторого множества элементарных событий. Но теперь степень принадлежности события $\{r_1, \dots, r_k\}$ к множеству возможных значений X , будет уже не *суммой* величин $p(r_1), \dots, p(r_k)$, а *максимальной* из них.

НВ X_1, \dots, X_k , характеризуемые функциями принадлежности соответственно $p_1(r), \dots, p_k(r)$, называются *независимыми*, если векторная НВ $Y = (X_1, \dots, X_k)$ такова, что степень принадлежности вектора (r_1, \dots, r_k) к множеству возможных значений Y равна *наименьшей* из величин $p_1(r_1), \dots, p_k(r_k)$.

Из приведенных определений, в частности, вытекает следующее правило *суммирования* независимых НВ: если НВ X_1 и X_2 характеризуются функциями принадлежности соответственно $p_1(r)$ и $p_2(r)$, то $Z = X_1 + X_2$ определяется как НВ, имеющая функцию принадлежности $p(r) = \sup_s \{ \max \{ p_1(s); p_2(r-s) \} \}$.

Частным случаем НВ являются величины, заданные интервально. Для них функция принадлежности равна 1 на некотором множестве (обычно – на интервале) возможных значений r и равна 0 вне его.

Приведем некоторые сведения из теории НВ [58]. Каждая НВ X однозначно характеризуется семейством своих "множеств уровня" (точнее – надуровня) $\mathcal{G}_X(h) = \{r | p_X(r) \geq h\}$, $0 < h \leq 1$. Множества этого семейства вложены друг в друга, но могут иметь произвольную форму¹⁸. Если все они выпуклы и замкнуты, назовем НВ *стандартной* (в одномерном случае стандартными будут НВ с унимодальными функциями принадлежности). Заменив каждое из множеств уровня своим выпуклым замыканием (в одномерном случае – наименьшим отрезком, содержащим множество уровня), мы получим другое семейство множеств, отвечающее новой, стандартной НВ X^c , которую мы назовем *выпуклым замыканием* исходной НВ X . В [58] предложено характеризовать близость НВ расстоянием между соответствующими множествами уровня – аналогично определялось расстояние между зависимостями в п. 2.1 и [6, 29].

При умножении НВ на положительное число λ множества уровня подобно преобразуются – каждая точка любого из этих множеств умно-

¹⁸ Семейство множеств уровня может быть также положено в основу определения понятия нечеткой величины. Так, нечеткую величину, можно определить как отображение отрезка $[0,1]$ в семейство вложенных друг в друга подмножеств числовой оси.

жается на λ . При сложении двух независимых НВ соответствующие множества уровня суммируются по Минковскому¹⁹.

Выясним, есть ли аналог закону больших чисел для НВ. Пусть S_n – *среднее арифметическое из n независимых НВ X_i с одинаковыми функциями принадлежности*. Из изложенных выше соображений несложно вывести, что если усредняемые НВ стандартны, то S_n является НВ с *той же самой функцией принадлежности*. Если же они не стандартны, то при $n \rightarrow \infty$ функция принадлежности для S_n стремится к функции принадлежности для выпуклого замыкания любой из X_i . Как видим, *аналога закону больших чисел для НВ нет – "случайные" колебания такого рода при усреднении взаимно не погашаются!*

Пусть Z – нечеткая ошибка наблюдения. Естественнее считать, что ее функция принадлежности $\pi = p_Z(r)$ велика при малых r и убывает с ростом $|r|$. Поэтому существует непрерывная обратная функция $|r| = g(\pi)$. Пусть $R = R_Z = g(1/2)$. Тогда любая ошибка, по модулю, не превышающая R , будет иметь большую степень принадлежности, чем любая ошибка, по модулю превышающая R . На этом основании величину R можно трактовать как *среднюю ошибку*.

Изложенная модель характерна для "обычных", т.е. "малых", ошибок. Для грубых ошибок ситуация несколько иная: здесь степень принадлежности тем больше, чем больше величина ошибки, поэтому функция g будет *возрастающей, а не убывающей*. Определение средней ошибки здесь сохраняется, но приобретает иной смысл: любая ошибка, по модулю не превышающая Δ , будет иметь *меньшую* степень принадлежности, чем любая ошибка, по модулю превышающая Δ .

Мы не будем останавливаться на технологии и методологии практического использования НВ, отсылая читателей к [56, 58 и др.]. В то же время представляется важным привести точку зрения Задэ, предложившего теорию нечетких множеств и дающего в предисловии к [59] такие объяснения целесообразности ее практического применения. "Первое объяснение может быть описано в терминах "мы не знаем", означая, что наше знание некоторой системы недостаточно точно или недостаточно полно для того, чтобы позволить нам использовать стандартные методы количественного анализа. Второе объяснение может быть описано в терминах "нам не так важно знать", что означает, что у нас нет необходимости знать какую-то систему с высокой степенью точности и детализации. Другими словами, нам не так важно, что информация неточна или частично недостоверна, если это может быть использовано для достижения хорошего и устойчивого (robust) решения с низкой стоимостью и хорошо согласованного с реальностью".

Второй альтернативой СВ являются введенные в [60] **величины, наделенные правдоподобием** (ПВ, *likelihood variables*). Эти величины

¹⁹ По Минковскому, сумма множеств \mathcal{A} и \mathcal{B} есть объединение всех точек вида $x + y$, где $x \in \mathcal{A}$, $y \in \mathcal{B}$.

задаются с помощью специальных функций правдоподобия так же, как устанавливаются вероятностные распределения для СВ или функции принадлежности для НВ (аксиоматически, экспертно или с помощью функций правдоподобия других величин).

Понятие степени правдоподобия при этом представляется более наглядным, чем понятие "степени принадлежности" для нечетких величин, формализуя представления субъекта о правдоподобности тех или иных событий и являясь основанием для проведения индуктивных рассуждений при неполной информации [61].

Как и НВ, ПВ X однозначно характеризуется некоторой функцией (относительного) правдоподобия $p_X(r)$, определенной на всем множестве возможных значений X , удовлетворяющей требованиям: $p_X(r) \geq 0$; $\sup p_X(r) = 1^{20}$. Числу $p_X(r)$ при этом задается смысл "правдоподобия r как возможного значения X ", точнее – как относительного правдоподобия r по сравнению с наиболее правдоподобным из возможных значений X . ПВ X отождествляется с детерминированным вектором r ; если $p_X(r) = 1$, $p_X(s) = 0$ при всех $s \neq r$. Легко убедиться, что функциям правдоподобия, равным 1 на некотором множестве (например, отрезке) и 0 вне его, отвечают интервально заданные величины.

Функции от ПВ определяются точно так же, как и для НВ, однако независимость ПВ вводится по аналогии с СВ. ПВ X_1, \dots, X_k , характеризуемые функциями правдоподобия соответственно $p_1(r), \dots, p_k(r)$, называются *независимыми*, если функция правдоподобия векторной ПВ $Y = (X_1, \dots, X_k)$ имеет вид $p_Y(r_1, \dots, r_k) = p_1(r_1) \dots p_k(r_k)$.

Отсюда, в частности, вытекает правило суммирования независимых ПВ: $Z = X_1 + X_2$ определяется как ПВ, имеющая функцию правдоподобия $p(r) = \sup_s [p_1(s)p_2(r-s)]$. В [57, 62] данное правило

вводилось как одно из возможных обобщений правила суммирования нечетких величин (однако ПВ нельзя рассматривать как разновидность НВ – эти величины выражают разные, но не все возможные виды неопределенности).

Приведем, аналогично вышесказанному, некоторые сведения из теории ПВ [60]. Каждая ПВ X однозначно характеризуется своим "логарифмическим надграфиком" $\mathcal{G}_X = \{(r, h) | -\ln p_X(r) \geq h\}$, $0 \leq h \leq 1$. В том случае, если это множество выпукло и замкнуто, назовем ПВ *стандартной* (в одномерном случае X будет стандартной, если функция $\ln p_X(r)$ выпукла). Заменяв логарифмический надграфик его выпуклым замыканием, мы получим другое множество, которое будет также логарифмическим надграфиком некоторой стандартной ПВ X^c – *выпуклого замыкания исходной ПВ X* .

Близость ПВ оказалось удобным характеризовать хаусдорфовым

²⁰ Последнее требование можно заменить двумя другими, возможно, более естественными: 1) функция p_X ограничена сверху; 2) функции p_X и kp_X при $k > 0$ определяют одну и ту же ПВ.

расстоянием между их логарифмическими надграфиками. При умножении ПВ на положительное число λ логарифмический надграфик подобно преобразуются – каждая его точка умножается на λ . При сложении двух независимых ПВ логарифмические надграфики суммируются по Минковскому. *Аналога закону больших чисел нет и для ПВ*: функция правдоподобия для среднего арифметического из n независимых "одинаково распределенных" ПВ X_i при $n \rightarrow \infty$ стремится к функции правдоподобия для выпуклого замыкания X_i (это доказывается так же, как и для НВ).

Как и для НВ, функции правдоподобия для ошибок наблюдений можно считать зависящими от модуля ошибки. Поэтому и здесь можно ввести понятие средней ошибки, отделяющей более правдоподобные (т.е. имеющие степень правдоподобия больше $1/2$) значения ошибок от менее правдоподобных. Так, если функция правдоподобия ошибки имеет функцию правдоподобия $2^{-(r/R)^2}$, то средняя ошибка будет равна R .

Будем называть функцию p_X функцией реальности независимо от того, какой из трех типов неопределенности она отражает.

Не следует думать, что все возможные типы неопределенных величин сводятся к трем описанным – их намного больше, и проблема полного описания класса "разумных" и практически важных типов неопределенности еще ждет своего решения. Отметим лишь, что "альтернативные" определения независимости должны, по нашему мнению, удовлетворять следующему требованию: если независимыми являются компоненты векторов $Z = (X_1, X_2)$ и $W = (Z, X_3)$, то независимыми будут и компоненты векторов $U = (X_1, X_3)$ и $V = (U, X_2)$. Такому требованию удовлетворяют не только операции умножения "степеней реальности" или взятия минимальной из них, но и другие, более "экзотические", например, $p_Z(r_1, \dots, r_n) = \max \left\{ 1, \sum_{i=1}^n p_{X_i}(r_i) \right\}$.

В то же время рассмотренные выше способы формализации "случайных" ошибок в определенном смысле могут рассматриваться как элементарные. Используя их как "кирпичики", можно получить "смешанные" типы неопределенности (см. п. 4.4), моделями которых удобно пользоваться при решении эконометрических задач. В этой связи в следующей таблице сопоставляются определения и основные свойства рассмотренных выше типов неопределенных скалярных величин.

Общую схему оптимальной оценки рассмотрим на примере оценки (векторного) параметра a в зависимости $Y = \Phi(X, a)$. Пусть Y_i и X_i – "истинные" значения характеристики i -го объекта ГС, X_i наблюдается точно, V_i – наблюдаемые (с ошибками) значения Y_i , а Y и V – вектора, образованные из Y_i и V_i для всех i . Если результатами наблюдений являются неопределенные величины одного из рассмотренных выше видов, то для каждого i при заданных Y_i наблюдение V_i характеризуется некоторой "степенью реальности" $\pi_i = p_i(Y_i, V_i, a)$. При этом

Основные типы неопределенных величин

Понятия, связанные с неопределенной величиной X	Случайная величина	Нечеткая величина	Величина, наделенная правдоподобием
Основная характеристика $p_X(r)$ – "степень реальности" события $X = r$	Вероятность события $X = r$	Степень принадлежности r к множеству возможных значений X	Правдоподобие r как возможного значения X
Свойства функции $p_X(r)$	$p_X(r) \geq 0,$ $\sum_r p_X(r) = 1$	$p_X(r) \geq 0,$ $\sup_r p_X(r) = 1$	$p_X(r) \geq 0,$ $\sup_r p_X(r) = 1$
X – детерминированная величина r	$p_X(r) = 1,$ $p_X(s) = 0; \quad \forall s \neq r$	$p_X(r) = 1,$ $p_X(s) = 0; \quad \forall s \neq r$	$p_X(r) = 1,$ $p_X(s) = 0; \quad \forall s \neq r$
$Y = f(X)$	$p_Y(s) = \sum_{r: f(r)=s} p_X(r)$	$p_Y(s) = \sup_{r: f(r)=s} p_X(r)$	$p_Y(s) = \sup_{r: f(r)=s} p_X(r)$
Независимость компонент вектора $Z = (X_1, \dots, X_n)$	$p_Z(r_1, \dots, r_n) = \prod_{i=1}^n p_{X_i}(r_i)$	$p_Z(r_1, \dots, r_n) = \min_i p_{X_i}(r_i)$	$p_Z(r_1, \dots, r_n) = \prod_{i=1}^n p_{X_i}(r_i)$

если сочетание (Y, V, a) в принципе невозможно (например, какие-то из показателей выходят за допустимые пределы), соответствующая степень реальности принимается равной нулю. Теперь по приведенным в таблице формулам можно определить и степень реальности для всей совокупности наблюдений $P(Y, V, a)$. Так, для независимых случайных наблюдений она равна произведению всех π_r , а для независимых нечетких наблюдений – наименьшему из π_r . При фиксированных результатах наблюдений V эта функция, как отмечено выше, будет отражать степень согласованности набора (Y, a) с данными наблюдений и другой исходной информацией, поэтому ее можно использовать как критерий согласованности (поскольку его значение лежит между нулем и единицей, он не будет инвариантен относительно монотонного преобразования). МС-принцип приводит теперь к задаче максимизации $P(Y, V, a)$ по переменным Y и a .

С формальной математической точки зрения методы оценки зависимостей при каждом из рассмотренных типов "случайных" ошибок идентичны: задается характер неопределенности результатов наблюдений и соответствующие функции реальности (они могут строиться на базе как объективной, так и субъективной информации, аналогично тому, как в статистике вводится функция нормального распределения ошибок), что позволяет сначала записать критерий согласованности имеющихся наблюдений с исходной информацией, а затем и выбрать

такую оценку зависимости, которая максимизирует этот критерий (соответствующие примеры приводятся ниже). Однако вероятностная формализация "случайных" ошибок заставляет дополнительно требовать повторяемости, воспроизводимости наблюдений, рассматривать их как выборку из некоторой теоретически бесконечной ГС. В то же время при описании неопределенности в терминах НВ или ПВ такие интерпретации не требуются – этими объектами могут быть описаны и неповторяющиеся, уникальные наблюдения. Естественной платой за это является отсутствие предельных теорем и несостоятельность оценок ("наиболее согласованные" значения оцениваемых параметров могут не сходиться к истинным при увеличении числа наблюдений), хотя при малых объемах наблюдений эту цену нельзя признать слишком высокой.

3.4. Интервальные неопределенности

Типичным примером неопределенности является следующий. Результаты наблюдения показывают, что величина X лежит в некотором интервале $[a, b]$. Неизвестен ни закон ее распределения, ни то, является ли она случайной величиной. В общем случае о неопределенной (например, векторной) характеристике X известно только ограниченное множество ее возможных значений. Этот вид неопределенности уже упоминался в п. 2.9.

Близкая ситуация рассматривалась в свое время Сэвиджем [49, 50], поставившим задачу о сравнении альтернатив, результаты которых зависят от неизвестного "состояния природы". При определенных предположениях он показал, что разумные правила сравнения при этом должны базироваться на введении "субъективных вероятностей" в пространстве возможных состояний природы. Приняв подход Сэвиджа, исследователь, казалось бы, может использовать "обычные" методы математической статистики, предварительно задав (экспертно) вероятностное распределение на указанном отрезке. Однако при этом придется убедительно (для "заказчика" и возможных рецензентов) обосновать правильность и объективность выбора именно этого распределения. С другой стороны, данную неопределенную величину X можно описать и как нечеткую, у которой функция принадлежности равна 1 на отрезке $[a, b]$ и нулю вне его, либо как ПВ с такой же функцией относительного правдоподобия. При этом имеющаяся информация о возможных значениях величины X используется полностью. Этот пример показывает, что стремление использовать "общепринятый" метод может усилить субъективизм получаемых при исследовании результатов.

Рассматриваемый тип "интервальной" неопределенности (set-uncertainty) допускает интересное обобщение [63, 64]. Если о величине X известно только, что она лежит в некотором интервале, то можно рассматривать ее как СВ, для распределения которой известен только носитель. Класс таких СВ весьма широк, но при наличии дополнительной

информации (например, о том, что правая половина интервала более вероятна, чем левая) он сужается. В результате возникает "интервально-вероятностный" тип неопределенности, когда ошибки (или результаты наблюдения) являются СВ, чей закон распределения принадлежит заданному классу \mathcal{F} (некоторому "интервалу" в пространстве распределений). В ситуации, когда \mathcal{F} состоит из единственного распределения, мы приходим к стохастическим величинам. Если \mathcal{F} состоит из распределений, зависящим известным образом от неизвестных параметров (например, из треугольных распределений с неизвестной вершиной или из усеченных нормальных распределений с неизвестным средним и дисперсией), возникает традиционная задача математической статистики: одновременное определение параметров неизвестной зависимости и закона распределения ошибок. Если же класс \mathcal{F} состоит из *всех* распределений на некотором отрезке, мы приходим к традиционной интервальной неопределенности. Наконец, если этот класс состоит из смесей вида $(1 - \epsilon)F + \epsilon H$, где F – распределение заданного типа, H – произвольное вероятностное распределение, мы приходим к модели грубых ошибок, рассмотренной в [42] и обсуждаемой ниже, в п. 4.3.

Подобная "смешанная" неопределенность может быть введена не только для СВ, но и для НВ и ПВ. Так, интервально-нечеткая (наделенная интервальным правдоподобием) величина может быть определена следующим образом. О величине X известно, что она нечеткая (наделена правдоподобием) и ее функция принадлежности (правдоподобия) принадлежит некоторому множеству \mathcal{F} . "Традиционная" интервальная неопределенность отвечает в этой ситуации множеству \mathcal{F} , включающему *любые* функции, значения которых на заданном отрезке лежат между 0 и 1 и равны 0 вне его.

4. ОЦЕНКА ЗАВИСИМОСТЕЙ ПРИ РАЗНЫХ ТИПАХ НЕОПРЕДЕЛЕННОСТИ

Если ничто другое не помогает, прочтите, наконец, инструкцию.

Аксиома Кана

4.1. Оценка параметрических зависимостей

В ситуации, когда результаты наблюдений являются НВ или ПВ, МС-принцип позволяет обосновать известные и ввести новые показатели интегрального отклонения и основанные на них критерии оптимальной оценки. Приведем ряд примеров.

Пример 4.1. Оценивается параметр a в зависимости $Y = \Phi(X, a)$ связывающей истинные значения показателей некоторой ГС объектов. При этом истинные значения X_i измеряются точно, а вместо истинных значений Y_i наблюдаются значения НВ $Y_i = Y_i + \xi_i$, отличающиеся от истинных аддитивными независимыми нечеткими ошибками ξ_i с одной и той же функцией принадлежности вида $h(|\xi_i|)$, где функция h монотонно убывает на положительной полуоси, $h(0) = 1$, $h(\eta) \rightarrow 0$ при $\eta \rightarrow +\infty$.

Тогда, поскольку величины X_i измеряются точно, степень принадлежности для всей совокупности значений (X_i, V_i) равна $\min_i h(|V_i - \Phi(X_i, a)|)$. Руководствуясь МС-принципом, оптимальное a следует выбрать из условия максимума этой функции. В силу свойств функции h этот максимум достигается, если (независимо от того, известна ли функция h точно или нет) параметр a является решением задачи

$$\max_i |V_i - \Phi(X_i, a)| \Rightarrow \min, \quad (4.1)$$

т.е. обеспечивает минимум минимального по модулю отклонения. Мы пришли к тому же минимаксному критерию, описанному в примере 2.5, однако уже при другом (не стохастическом) характере ошибок наблюдения! В простейшем случае, когда функция Φ вообще не зависит от X , а искомая зависимость имеет вид $Y_i = a$ (задача нахождения среднего значения), оптимальное решение будет средним арифметическим из наибольшего и наименьшего наблюдаемых значений показателя. ■

В п. 2.7 мы обращали внимание на неустойчивость МП-оценок. Рассмотренный пример показывает, что в условиях, когда ошибки наблюдения являются НВ, МС-метод приводит к устойчивым оценкам: небольшие изменения вида функции h не сказываются на оптимальной оценке вообще. Если же допустить, что истинная функция принадлеж-

ности мало отличается от h , но, в отличие от h , несимметрична, оценка изменится, но незначительно. Достаточно устойчивыми оказываются оценки и тогда, когда ошибки наблюдений являются ПВ (см. ниже). Казалось бы, структура критерия согласованности здесь та же, что и при случайных ошибках. Однако здесь есть и принципиальное различие. В стохастическом случае функция правдоподобия не допускает однозначной трактовки: в одних случаях она понимается как *вероятность* наблюдения тех или иных значений переменных, в других – как *плотность* распределения этих вероятностей. Если же рассматривать значения переменных как ПВ, то функция правдоподобия всегда имеет один и тот же смысл, выражая степень правдоподобия того или иного значения показателя. При этом не так уж важно, возникла ли эта функция из самих вероятностей, их плотностей или каким-либо иным образом. С этой позиции теория вероятностей должна рассматриваться в эконометрике лишь как удобная в определенных ситуациях "платформа" для построения функций согласованности оценок с данными наблюдений, но отнюдь не как обоснование стохастичности самих этих данных или ошибок в них.

В п. 2.6 мы отмечали, что при наличии случайных ошибок в нескольких наблюдаемых переменных функция правдоподобия может трактоваться двояко. Однако, если ошибки являются НВ или ПВ, подобной двусмысленности не возникает.

Пусть, например, оцениваются параметры в зависимости $Y = \Phi(X, Z, a)$, связывающей истинные значения показателей некоторой ГС объектов. При этом истинные значения показателей Z_i и Y_i наблюдаются точно, а вместо истинных значений X_i наблюдаются значения НВ $U_i = X_i + \xi_i$, отличающиеся от истинных аддитивными независимыми нечеткими (векторными) ошибками ξ_i с одной и той же функцией принадлежности вида $h(\xi)$. Таким образом, здесь должны выполняться соотношения:

$$Y_i = \Phi(U_i - \xi_i, Z_i, a). \quad (4.2)$$

Трактуя функцию согласованности здесь в духе МПО-метода (п. 2.6), мы должны поставить задачу отыскания оптимальных ошибок ξ_i . При этом степень реальности оптимального сочетания величин ξ_i будет равна $\min_i h(\xi_i)$, так что оптимальное a максимизирует эту функцию при ограничениях (4.2):

$$\begin{aligned} a &= \arg \sup \left\{ \min_i h(\xi_i) \mid Y_i = \Phi(U_i - \xi_i, Z_i, a) \forall i \right\} = \\ &= \arg \sup \left\{ \min_i h(U_i - X_i) \mid Y_i = \Phi(X_i, Z_i, a) \forall i \right\}. \end{aligned} \quad (4.3)$$

При МПЗ-методе функция согласованности отражает степень принадлежности совокупности нечетких величин $f_i = \Phi(U_i - \xi_i, Z_i, a) - Y_i$. Очевидно, что она равна верхней грани величин $h(\xi_i)$, взятой по всем их

сочетаниям, обеспечивающим заданные значения f_i . Поэтому одновременному обращению в нуль всех f_i отвечает степень правдоподобия, равная $\min_i \sup \{h(\xi_i) \mid Y_i = \Phi(U_i - \xi_i, Z_i, a)\}$, так что здесь

$$\begin{aligned} a &= \arg \min_i \{ \sup h(\xi_i) \mid Y_i = \Phi(U_i - \xi_i, Z_i, a) \} = \\ &= \arg \min_i \{ \sup h(U_i - X_i) \mid Y_i = \Phi(U_i, Z_i, a) \}. \end{aligned} \quad (4.4)$$

Нетрудно убедиться, что в данной ситуации минимум по t и супремум по ξ_i перестановочны, так что (4.3) и (4.4) дают одни и те же результаты.

В качестве $h(\xi)$ естественно использовать функции, убывающие по мере удаления точки ξ от начала координат, т.е. зависящие от некоторой нормы вектора ξ : $h(\xi) = f(\|\xi\|)$, где функция f монотонно убывает на положительной полуоси, $f(0) = 1$, $f(\eta) \rightarrow 0$ при $\eta \rightarrow +\infty$. В такой ситуации (независимо от того, известна ли функция f точно или нет) параметр a обеспечивает минимум максимального из расстояний от точек U_i до соответствующих поверхностей $Y_i = \Phi(X_i, Z_i, a)$. В частности, когда переменные Z в искомой модели вообще отсутствуют, эти поверхности становятся изоквантами функции Φ и мы приходим к методу измерения локальных отклонений по отклонениям объясняющих переменных от изоквант, описанному в п. 2.3.

Пример 4.2. Оцениваются параметры a_i и b_k зависимости $Y = \sum_{i=1}^m a_i X_i + \sum_{k=1}^n b_k Z_k$. В t -м наблюдении значения Z_{kt} и Y_t наблюдаются точно, а вместо истинных значений X_{it} наблюдаются значения НВ $U_{it} = X_{it} + \xi_{it}$. Ошибки ξ_{it} при разных i и t независимы и имеют функцию принадлежности $h(|\xi_{it}|/S_i)$, где S_i — известные масштабные параметры, характеризующие средние размеры ошибок, h — убывающая функция.

Здесь оптимальные оценки неизвестных параметров являются решением задачи:

$$\min_t \left\{ \sup_i \min h(|\xi_{it}|/S_i) \mid Y_t = \sum_{i=1}^m a_i (U_{it} + \xi_{it}) + \sum_{k=1}^n b_k Z_{kt} \right\} \Rightarrow \max.$$

Легко проверяется, что при каждом t оптимальные ξ_{it} имеют одну и ту же степень принадлежности, так что

$$\xi_{it} = \frac{\sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t}{\sum_{i=1}^m S_i |a_i|} S_i \operatorname{sign}(a_i).$$

Поскольку функция h убывающая, отсюда следует, что

оптимальные a_i и b_k являются решением задачи:

$$\frac{\max_t \left| \sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t \right|}{\sum_{i=1}^m S_i |a_i|} \Rightarrow \min. \blacksquare \quad (4.5)$$

Пример 4.3. Пусть в условиях примера 4.2 независимые ошибки наблюдения ξ_{it} являются ПВ и характеризуются функциями правдоподобия $h(\xi_{it}) = \exp\{-\xi_{it}^2/2S_i\}$. Тогда степень правдоподобия всей совокупности ошибок будет $\exp\left\{-\sum_t \left[\sum_{i=1}^m \xi_{it}^2 / 2S_i \right]\right\}$ и задача сведется к максимизации этого выражения при условиях $Y_t = \sum_{i=1}^m a_i (U_{it} - \xi_{it}) + \sum_{k=1}^n b_k Z_{kt}$.

Отсюда найдем: $\xi_{it} = a_i S_i \frac{\sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t}{\sum_{i=1}^m a_i^2 S_i}$. Из этого легко выводится, что оптимальные a_i и b_k являются решением задачи:

$$\frac{\sum_t \left\{ \sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t \right\}^2}{\sum_{i=1}^m a_i^2 S_i} \Rightarrow \min. \quad (4.6)$$

Так, если $n = 0$, а все S_i одинаковы, оптимальным a_i отвечает минимальная сумма квадратов расстояний от точек (U_{1t}, \dots, U_{mt}) до гиперплоскостей $Y_t = \sum_{i=1}^m a_i X_i$. Такой критерий был, по-видимому, впервые предложен в [65]. ■

Пример 4.4. Оценивается параметр \mathbf{a} в зависимости $Y = \Phi(X, \mathbf{a})$, причем значения X_t наблюдаются точно, а результатом наблюдения Y_t является $V_t = Y_t + \xi_t$, где ξ_t – независимые ПВ, имеющие одну и ту же функцию относительного правдоподобия $h(\xi)$. Здесь МС-принцип приводит к задаче: $\prod_t h[V_t - \Phi(X_t, \mathbf{a})] \Rightarrow \max$. Легко видеть, что такой критерий эквивалентен минимизации показателя интегрального отклонения $\sum_t g[V_t - \Phi(X_t, \mathbf{a})]$, где $g(x) = -\ln h(x)$. Критерии такого типа предложены в [42] для устойчивой оценки параметров вероятностных распределений. Отметим важное их преимущество: изменения функции h , при которых ее логарифмическая производная меняется мало, несущественно влияют на оптимальное значение \mathbf{a} . ■

Обратим теперь внимание на то, что при данном подходе оптимальные оценки параметров минимизируют показатель интегрального отклонения, исчисленный тем или иным способом (в примере 4.4 – сумма некоторых функций от локальных отклонений, в примере 4.1 – максимум из локальных отклонений). Это не противоречит общей оценке метода минимизации интегрального отклонения, высказанной в п. 2.3, поскольку в рассмотренных ситуациях неопределенность наблюдений характеризовалась определенной структурой – для функций правдоподобия иного вида окончательный критерий может не сводиться к расстояниям, измеряемым в легко интерпретируемых метриках.

Пример 4.5. Пусть в условиях примера 4.3 функция правдоподобия ошибок имеет иной вид $h(\xi_{it}) = \frac{1}{\text{ch}(\xi_{it} / D_i)}$. Легко проверить, что опти-

мизация ξ_{it} сводится здесь к задаче: $\sum_{i=1}^m \ln \text{ch}(\xi_{it} / D_i) \Rightarrow \min$;

$$\sum_{i=1}^m a_i \xi_{it} = \sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t = \Delta_t.$$

Поэтому $\xi_{it} = D_i \text{arth}(\lambda_i a_i D_i)$, где λ_i – корень уравнения $\sum_{i=1}^m a_i D_i \text{arth}(\lambda_i a_i D_i) = \Delta_t$. При этом $\ln \text{ch}(\xi_{it} / D_i) = -\frac{1}{2} \ln[1 - (\lambda_i a_i D_i)^2]$, так что оптимальные a_i и b_k являются решением задачи:

$$\sum_{i=1}^m \ln \text{ch}[1 - (\lambda_i a_i D_i)^2] \Rightarrow \min;$$

$$\sum_{i=1}^m a_i D_i \text{arth}(\lambda_i a_i D_i) = \sum_{i=1}^m a_i U_{it} + \sum_{k=1}^n b_k Z_{kt} - Y_t \quad \forall t.$$

Такой критерий уже не сводится к минимизации какого-то интегрального отклонения.■

Мы видим, что минимизация интегрального отклонения как метод оптимальной оценки параметров зависимостей обоснована лишь в случаях, когда "степень реальности" наблюдений определяется их расстояниями от истинных значений характеристик – это довольно распространенная, но отнюдь не общая ситуация.

Отметим, наконец, что МС-принцип применим и при неточно заданных функциях правдоподобия. Пусть в условиях примера 4.4 о функции правдоподобия h известен лишь некоторый класс \mathcal{H} , которому она принадлежит. Тогда задача сведется к одновременному нахождению наиболее согласованных с исходной информацией значений параметра α и функции $h \in \mathcal{H}$ из условия $\prod_t h\{V_t - \Phi(X_t, \alpha)\} \Rightarrow \max$.

4.2. Оценка непараметрических зависимостей

До сих пор речь шла, в основном, об оценке параметрических зависимостей. Как доказывается в математической статистике [2], если в таких зависимостях объясняемая переменная наблюдается со случайной ошибкой, то МП-оценки неизвестных параметров оказываются состоятельными, т.е. с вероятностью 1 сходятся к истинным значениям при неограниченном увеличении числа наблюдений. Для непараметрических зависимостей положение иное. Действительно, пусть, например, вид зависимости $Y = f(X)$ функции неизвестен, значения Y наблюдаются только в точках $X = 0, 1, 2$, и любая другая информация о моделируемом процессе или явлении отсутствует. Тогда никакое количество наблюдений не позволит даже приближенно оценить значение $f(3)$. Это означает, что для непараметрических зависимостей рассчитывать на получение состоятельных оценок в общем случае не приходится и увеличение числа наблюдений не позволит, в общем случае, повысить точность оценки.

Рассмотрим в этой связи задачу оценки неизвестной зависимости $Y = f(X)$ по результатам T наблюдений в точках $(X_1 < X_2 < \dots < X_T)$. Предполагается, что значения X_i измеряются точно, а вместо истинных значений $Y_i = f(X_i)$ наблюдаются значения $U_i = Y_i + \xi_i$, отличающиеся от истинных аддитивными независимыми случайными ошибками ξ_i , имеющими нормальное распределение с нулевым средним и неизвестной дисперсией S . Естественно, что такой информации явно недостаточно для получения разумных решений – необходимы дополнительные сведения о характере оцениваемой зависимости. Исследуем два варианта такой дополнительной информации.

Пример 4.6. Пусть известно, что искомая функция $f(X)$ – *неубывающая*. Этого, разумеется, недостаточно для того, чтобы установить характер изменения функции **между** точками X_i или за пределами отрезка $[X_1, X_T]$. Попробуем, однако, оценить значения неизвестной функции $f(X)$ в точках X_i . Функция правдоподобия в данном случае равна $\prod_i \frac{1}{\sqrt{2\pi S}} \exp\{-[U_i - f(X_i)]^2 / 2S\} = \prod_i \frac{1}{\sqrt{2\pi S}} \exp\{-[U_i - Y_i]^2 / 2S\}$, поэтому ее максимизация эквивалентна решению задачи:

$$\sum_i (Y_i - U_i)^2 \Rightarrow \min, \quad Y_1 \leq Y_2 \leq \dots \leq Y_T. \quad (4.7)$$

Решив эту задачу выпуклого программирования известными численными методами, можно оценить искомую функцию и в "промежуточных" точках. Действительно, ее значения между точками X_{i-1} и X_i лежат, очевидно, в пределах от Y_{i-1} до Y_i . Поэтому разумной оценкой для $f(X)$ будет, например, ломаная, соединяющая точки (X_i, Y_i) . Выше подразумевалось, что все точки X_i – различные. Однако решение принципиально не изменится, если считать, что некоторые наблюдения про-

изведены при одинаковых значениях объясняющей переменной – в этой ситуации в качестве неизвестных можно принять только значения Y_t , отвечающие разным X_t , либо добавить дополнительные ограничения $Y_t = Y_{t-1}$ для тех t , для которых $X_t = X_{t-1}$. ■

Пример 4.7. Пусть известно, что искомая функция $f(X)$ – выпуклая. Рассуждая аналогично, мы приходим к другой задаче выпуклого программирования:

$$\sum_t (Y_t - U_t)^2 \Rightarrow \min, \quad \frac{Y_2 - Y_1}{X_2 - X_1} \leq \dots \leq \frac{Y_T - Y_{T-1}}{X_T - X_{T-1}}. \quad (4.8)$$

Тот же метод может быть применен и при оценке функций $Y = f(X)$ нескольких переменных, хотя формализация модели здесь гораздо более сложная. При этом, как и в одномерном случае, удастся оценить функцию *только в ограниченной области*, являющейся выпуклой оболочкой наблюдаемых точек X_i . ■

Пример 4.8. Требуется оценить **плотность распределения** $f(X)$ случайной величины по T наблюдаемым ее значениям X_i . Как и в предыдущих случаях, одной этой информации (и даже дополнительной информации о том, что эта плотность "колоколообразна"), недостаточно. Однако для типичных распределений такого вида функция $g(X) = -\ln f(X)$ выпуклая (мы допускаем и такие выпуклые функции, которые принимают значение $+\infty$). Оказывается, что эта информация уже позволяет получить разумные оценки. Действительно, логарифм функции правдоподобия наблюдаемых значений $\{X_i\}$, взятый с обратным знаком, равен сумме $\sum_i g(X_i)$, поэтому задача сводится к нахождению такой выпуклой функции $g(X)$, для которой

$$L(g) = \sum_{i=1}^T g(X_i) \Rightarrow \min \quad (4.9)$$

при естественном ограничении на плотность вероятностного распределения:

$$\int_{-\infty}^{\infty} e^{-g(X)} dX = 1. \quad (4.10)$$

Решение этой задачи оказывается сравнительно простым. Пусть H – минимальное значение $L(g)$, $h(X)$ – произвольная выпуклая функция,

$J(h) = \int_{-\infty}^{\infty} e^{-h(X)} dX$. Легко видеть, что функция $h(X) + \ln J(h)$ удовлетворяет (4.10), так что для нее левая часть (4.9) будет не меньше H .

Отсюда следует, что величина H является минимальным значением

функционала $R(h) = \sum_{i=1}^T h(X_i) + T \ln \left[\int_{-\infty}^{\infty} e^{-h(X)} dX \right]$ на множестве всех вы-

пуклых функций. Поэтому оптимальная функция $h(X)$ заведомо должна обращаться в $+\infty$ при $X < X_I$ или при $X > X_T$ (соответственно в этих точках будет $g(X) = 0$). Обозначим $Z_t = h(X_t)$; $U_t = \frac{Z_{t+1} - Z_t}{X_{t+1} - X_t}$. Тогда, очевидно,

$$U_I \leq U_2 \leq \dots \leq U_{T-1}, \quad (4.11)$$

$$Z_t = Z_I + \sum_{i=1}^{t-1} U_i (X_{i+1} - X_i) \quad (4.12)$$

Рассмотрим ломаную $h^*(X)$, соединяющую последовательные точки (X_t, Z_t) :

$$h^*(X) = Z_t + U_t(X - X_t) \quad \text{при} \quad X_t < X < X_{t+1}. \quad (4.13)$$

Поскольку $h(X)$ выпукла, то $h^*(X) \leq h(X)$ и, значит, $R(h) \leq R(h^*)$, что возможно только, если $h(X) = h^*(X)$. Проведя несложные вычисления, получим, что тогда

$$\begin{aligned} R(h) &= \sum_{t=1}^T Z_t + T \ln \left[\sum_{t=0}^{T-1} e^{-Z_t} \frac{1 - e^{-U_t(X_{t+1} - X_t)}}{U_t} \right] = \\ &= \sum_{t=1}^{T-1} (T-t) U_t (X_{t+1} - X_t) + T \ln \left[\sum_{t=1}^{T-1} \frac{1 - e^{-U_t(X_{t+1} - X_t)}}{U_t} e^{-\sum_{i=1}^{t-1} U_i (X_{i+1} - X_i)} \right]. \end{aligned} \quad (4.14)$$

Минимизация (4.14) при ограничениях (4.11) является задачей выпуклого программирования и легко решается с помощью известных компьютерных программ. Легко видеть, что всем ломаным $h(X)$ с оптимальными U_t будет отвечать одно и то же значение $R(h)$. Однако отрицательным логарифмом плотности распределения из них будет только одна – для нее должно выполняться равенство (4.10). Отсюда можно найти последнее неизвестное Z_I :

$$Z_I = \ln \left[\sum_{t=1}^{T-1} \frac{1 - e^{-U_t(X_{t+1} - X_t)}}{U_t} e^{-\sum_{i=1}^{t-1} U_i (X_{i+1} - X_i)} \right]. \quad (4.15)$$

Теперь оптимальная $g(X)$ определяется из (4.13). Приведем численные примеры.

Получены 12 реализаций случайной величины, равномерно распределенной на $(0,1)$:

0,012; 0,036; 0,229; 0,252; 0,325; 0,489; 0,491; 0,534; 0,555; 0,771;
0,943; 0,994.

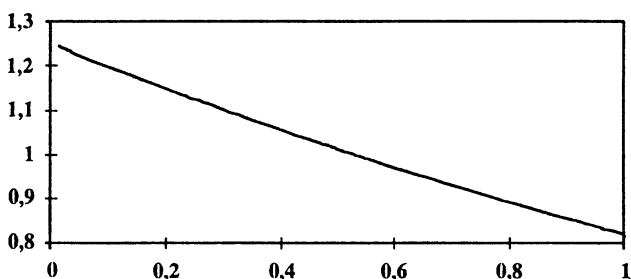


Рис. 1. Оценка плотности равномерного распределения

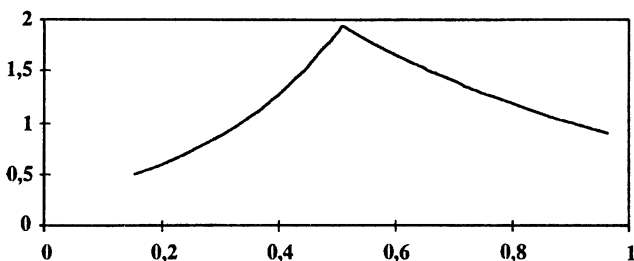


Рис. 2. Оценка плотности треугольного распределения

На рис. 1 приведена оцененная указанным способом плотность распределения.

В другом расчете получены 12 реализаций случайной величины, имеющей треугольное распределение на $(0,1)$:

0,153; 0,370; 0,391; 0,485; 0,510; 0,516; 0,629; 0,639; 0,662; 0,735; 0,895; 0,964.

Оцененная указанным способом плотность распределения приведена на рис. 2.

Если ошибки являются НВ или ПВ, оценка непараметрических зависимостей иногда может быть получена на основе МС-принципа. Приведем ряд примеров.

Пример 4.9. Оценивается функция $f(X)$ по результатам наблюдений $Y_i = f(X_i) + \xi_i$, где ξ_i – нечеткие ошибки, имеющие одну и ту же симметричную унимодальную функцию принадлежности $g(x)$. Будем считать, что априорная информация об искомой функции позволяет считать ее нечеткой. Иными словами, предположим, что каждой из априори возможных функций f отвечает некоторая степень реальности, в данном случае – степень принадлежности $Q(f)$, тем меньшая, чем менее "гладкой" будет f . Примем поэтому, что она имеет вид $H(\int |f'(X)|^2 dX)$, где H – некоторая убывающая функция, $H(0) = 1$, $H(\infty) = 0$. Считая, что функция f и ошибка ξ , независимы, получим значение функции

принадлежности для их совместной реализации:

$$\begin{aligned} & \min \left\{ H(\int |f''(X)|^2 dx), \min_i g(\xi_i) \right\} = \\ & = \min \left\{ H(\int |f''(X)|^2 dx), \min_i g(|Y_i - f(X_i)|) \right\}. \end{aligned}$$

Максимизация этого критерия согласованности эквивалентна решению задачи:

$$\int |f''(X)|^2 dx \Rightarrow \min, |Y_i - f(X_i)| \leq D, \quad (4.16)$$

при некотором D , зависящем от g и H . Легко проверяется, что решением этой задачи будет некоторый сплайн. В п. 2.8 такая формализация задачи была получена из эвристических соображений. ■

Аналогично может быть поставлена и задача оценки функции $f(X)$ при наличии нечетких ошибок наблюдений как объясняемой, так и объясняющих переменных. Ситуация, когда ошибки являются ПВ, рассматривается в следующем примере.

Пример 4.10. Оценивается функция $f(X)$ по результатам наблюдений $Y_i = f(X_i) + \xi_i$, где ξ_i – ошибки, характеризуемые функцией относительного правдоподобия $e^{-k\xi^2}$. Будем считать, что априорная информация об искомой функции позволяет считать ее наделенной правдоподобием, причем степень относительного правдоподобия каждой из априори возможных функций f зависит от ее "гладкости" и имеет вид $\exp\{-\int |f''(X)|^2 dx\}$. Считая, что функция f и ошибки ξ_i независимы, запишем отрицательный логарифм функции правдоподобия для их совместной реализации, который одновременно является и критерием оптимальной оценки f :

$$L = \int |f''(X)|^2 dx + k \sum_{i=1}^T [f(X_i) - Y_i]^2 \Rightarrow \min. \quad (4.17)$$

Подобный критерий из эвристических соображений приведен в п. 2.8.

Тот же метод можно использовать и для оценки более сложных зависимостей.

Пример 4.11. Оценивается производственная функция вида $V_t = \Phi(K_t, L_t)$, где V_t – валовой продукт, K_t – используемый капитал, L_t – затраты труда (например, количество отработанных человеко-часов) в году t . Предполагается, что оцениваемая функция – однородная первой степени. Прямое измерение величин V_t, K_t, L_t невозможно, известны лишь их наблюдения v_t, k_t, l_t , отличающиеся от истинных значений независимыми мультипликативными ПВ-ошибками:

$$\ln v_t = \ln V_t + \xi_{vt}; \quad \ln k_t = \ln K_t + \xi_{kt}; \quad \ln l_t = \ln L_t + \xi_{lt}.$$

Предполагается, что функции относительного правдоподобия для этих ошибок имеют вид e^{-bx^2} с разными параметрами масштаба b_v , b_k и b_l , одинаковыми для всех наблюдений ("равноточность" разных наблюдений).

Для решения этой задачи обозначим $\varphi(x) = \ln[\Phi(e^x, 1)]$ и перейдем к показателям (истинной и наблюдаемой) производительности и капиталоемкости труда:

$$P_t = \ln\left(\frac{V_t}{L_t}\right); \quad p_t = \ln\left(\frac{v_t}{l_t}\right); \quad Q_t = \ln\left(\frac{K_t}{L_t}\right); \quad q_t = \ln\left(\frac{k_t}{l_t}\right);$$

Тогда истинная зависимость и соотношения между истинными и наблюдаемыми переменными примут вид: $P_t = \varphi(Q_t)$, $p_t = P_t + \xi_{vt} - \xi_{lt}$; $q_t = Q_t + \xi_{kt} - \xi_{lt}$, откуда

$$\xi_{kt} = q_t - Q_t + \xi_{lt}; \quad \xi_{vt} = p_t - \varphi(Q_t) + \xi_{lt}. \quad (4.18)$$

Если бы функция φ была известна, согласованность с ней результатов наблюдений можно было бы охарактеризовать функцией относительного правдоподобия $\prod_t \text{Пехр}\{-b_v \xi_{vt}^2 - b_k \xi_{kt}^2 - b_l \xi_{lt}^2\}$. Однако функция φ неизвестна, но имеющаяся информация позволяет считать ее достаточно гладкой. Примем поэтому, что она наделена правдоподобием и соответствующая функция правдоподобия имеет вид $\exp\{-\lambda \int |\varphi''(x)|^2 dx\}$. Тогда критерий согласованности, учитывающий одновременно и гладкость и распределение ошибок наблюдения, примет вид:

$$e^{-\lambda \int |\varphi''(x)|^2 dx} \times \prod_t \text{Пехр}\{-b_v \xi_{vt}^2 - b_k \xi_{kt}^2 - b_l \xi_{lt}^2\} \Rightarrow \max.$$

Используя (4.18), величины ξ_{vt} и ξ_{kt} можно выразить через ξ_{lt} и Q_t , что позволяет представить критерий оптимальности в более простом виде:

$$\lambda \int |\varphi''(x)|^2 dx + \sum_t \{b_v [p_t - \varphi(Q_t) + \xi_{lt}]^2 + b_k (q_t - Q_t + \xi_{lt})^2 + b_l \xi_{lt}^2\} \Rightarrow \min.$$

Минимизируя по переменным ξ_{lt} , приходим к задаче нахождения оптимальной функции φ и неизвестных Q_t по критерию:

$$\lambda \int |\varphi''(x)|^2 dx + \sum_t \left\{ b_v [p_t - \varphi(Q_t)]^2 + b_k (q_t - Q_t)^2 - \frac{|b_v [p_t - \varphi(Q_t)] + b_k (q_t - Q_t)|^2}{b_v + b_k + b_l} \right\} \Rightarrow \min.$$

Решением ее будет некоторый сплайн, однако "точки склеивания" его полиномиальных участков определяются нелинейными соотношениями, так что данную задачу легче решать численными методами. ■

Пример 4.12. Известны результаты X_i наблюдений случайной величины. Требуется оценить **плотность** $f(X)$ ее распределения. В примере 4.8 эта задача решалась МП-методом в предположении, что функция $-\ln f(X)$ выпуклая. МС-принцип позволяет решить эту задачу при более слабом предположении о *гладкости* $f(X)$. Такую информацию можно формализовать, рассмотрев плотность $f(X)$ как *неопределенную* функцию, наделенную правдоподобием, причем степень ее правдоподобия связана с функционалом гладкости. Поскольку плотность неотрицательна, введем в рассмотрение ее логарифм, обозначив $u(x) = \ln f(x)$. По аналогии с примером 4.10 примем, что априорная степень правдоподобия $u(X)$ равна $\exp\{-k|u''(X)|^2 dx\}$. Степень правдоподобия наблюдаемых значений $\{X_i\}$ при известной $u(X)$ в данном случае равна $\prod_{i=1}^T e^{u(X_i)}$. Поэтому критерий согласованности в данном случае

принимает вид: $\exp\left\{-k \int |u''(X)|^2 dX + \sum_{i=1}^T u(X_i)\right\}$, а оптимальная оценка будет решением задачи:

$$k \int |u''(X)|^2 dX + \sum_{i=1}^T u(X_i) \Rightarrow \min, \quad \int_{-\infty}^{\infty} e^{u(X)} dX = 1. \quad (4.19)$$

Аналитическое решение этой задачи затруднительно, однако известные компьютерные программы позволяют решать ее численно. ■

4.3. Выявление грубых ошибок и устранение их влияния

Некоторые из ошибок наблюдения рассматриваются на практике как "грубые". Один тип таких ошибок приводился выше – это ошибки спецификации, связанные с неверным в каком-то смысле выбором объекта наблюдения. Другой пример грубых ошибок часто возникает при сборе экономической информации, когда при ручном заполнении статистических таблиц искажается положение десятичной запятой. Ошибки такого рода часто удастся выявить визуально: при графическом представлении собранной информации соответствующие точки выглядят на графике как "резко выделяющиеся". Однако если ошибка допущена в нескольких показателях объекта, выявить ее значительно труднее. Вопросам обработки информации методами математической статистики при наличии грубых ошибок посвящены, например, работы [42, 44], а также отдельные разделы в [1]. Между тем понятие "грубых ошибок" допускает неоднозначные трактовки. Это связано с тем, что в отличие от "стандартного" подхода, рассматривающего совокупность

наблюдений как *однородную* (характеризующуюся некоторым единым законом распределения), здесь она считается *неоднородной*, объединяющей наблюдения, относящиеся к разным однородным совокупностям (например, к данным о строительных организациях добавлены и данные, относящиеся к предприятиям других отраслей). Однако дать строгое определение "неоднородности" совокупности затруднительно даже в случае вероятностей неопределенности. Действительно, пусть зависимость между Y и X имеет вид $Y = aX + \xi$, причем ошибки ξ имеют нормальное распределение с нулевым средним и дисперсией, пропорциональной X (см. примеры 2.4 и 2.9). Формально получаемая совокупность наблюдений (X, Y) неоднородна, ибо ошибки в разных наблюдениях имеют разные вероятностные распределения. С другой стороны, эту же совокупность можно считать однородной, представив ξ иначе, в виде $\xi = \eta X^{1/2}$ – теперь все "нормированные ошибки" η будут иметь одно и то же распределение.

При математико-статистическом подходе ГС рассматривают как неоднородную, если она может быть расщеплена на несколько классов (групп) с разными законами распределения (см. пример 2.8). Однако именно для грубых ошибок такой подход неудобен: здесь закон распределения грубых ошибок не является объектом оценки и информации о нем не сопоставима с информацией о распределении "обычных" ошибок. Поэтому данное определение целесообразно скорректировать: ГС считается неоднородной, если она расщепляется на несколько классов с *разной информацией* о распределении характеристик входящих в них объектов. Однако различие информации о характере неопределенности результатов "обычных" и "грубо ошибочных" наблюдений может быть формализовано по-разному. Поэтому возможны разные формализации (модели) "грубых ошибок" и соответственно – разные подходы к сепарации обычных и "грубых" ошибок. Переходя к их рассмотрению, предварительно заметим, что применительно к НВ и ПВ соответствующие задачи ранее не рассматривались. При этом в большинстве примеров мы рассматриваем **базовую** задачу оценки параметра a в зависимости $Y = \Phi(X, a)$ между истинными значениями характеристик некоторых объектов. Предполагается, что истинные значения X_i измеряются точно, а вместо истинных значений Y_i наблюдаются величины $V_i = Y_i + \xi_i$, отличающиеся от истинных аддитивными независимыми ошибками ξ_i , среди которых могут быть и грубые (разных типов). Количество грубых ошибок мы обозначаем через k , а множество номеров "грубо ошибочных наблюдений" – через \mathcal{K} .

4.3.1. Модель "засорения"

Исторически первой моделью "грубых ошибок" была, по-видимому, модель "засорения" Хьюбера [42], предложенная для оценки параметров вероятностных распределений. Здесь считаются известными наблюдения случайной величины, которая *должна* иметь распределение

F с плотностью $p(X, \Theta)$ с неизвестным параметром Θ . Однако реально такая величина наблюдается лишь с вероятностью $1-\epsilon$, а с дополнительной вероятностью наблюдается совсем другая случайная величина, имеющая другое, "засоряющее" вероятностное распределение H , не обязательно обладающее плотностью, о котором известно лишь, что оно принадлежит определенному широкому классу \mathcal{H} (например, что оно произвольно или центрально-симметрично). Распределение результатов наблюдений при этом оказывается *смесью* $(1-\epsilon)F + \epsilon H$. Метод оценки Θ при этом подбирается так, чтобы минимизировать максимальное (по всем $H \in \mathcal{H}$) ожидаемое влияние "засорения" (т.е. так, чтобы при наихудшем засорении вызванная им ошибка в оцениваемых параметрах была в среднем наименьшей). Обратим особое внимание, что МП-метод в этих целях не подходит, поскольку распределения вида $(1-\epsilon)F + \epsilon H$ могут не иметь плотности.

Легко видеть, однако, что предположение о стохастическом характере ошибок с методической точки зрения не обязательно. Действительно, изложенная модель базируется на предположении о том, что закон распределения результатов наблюдений в определенном смысле близок к заданному. Тот факт, что речь идет именно о вероятностном распределении, здесь принципиальной роли не играет. Поэтому в задачах установления зависимостей при наличии ошибок, являющихся НВ или ПВ, соответствующий вариант модели засорения будет выглядеть просто: *функции принадлежности или относительного правдоподобия ошибок близки в том или ином смысле к заданным*, например, отстоят от них на расстояние не большее ϵ .

В этой ситуации функция реальности ошибки известна неточно — известен лишь некоторый класс \mathcal{F} , которому она принадлежит (для "хьюберовского" засорения \mathcal{F} состоит из всех функций распределения, отличающихся от заданной не больше чем на ϵ в равномерной метрике). Поэтому, если ошибки наблюдений являются НВ или ПВ, каких-либо препятствий для применения данного метода нет.

Пример 4.13. Примем, что в базовой модели ошибки наблюдений ξ , являются ПВ. Как и в модели Хьюбера, будем считать, что класс \mathcal{F} состоит из всех функций правдоподобия, отличающихся от заданной функции g не более, чем на ϵ . Для этого, естественно, надо ввести метрику на классе функций правдоподобия. Воспользуемся метрикой [58], в которой расстояние ρ между функциями правдоподобия определяется как максимальное хаусдорфово расстояние между соответствующими множествами их уровня (см. п. 3.3). Тогда класс \mathcal{F} состоит из функций h таких, что при любом d множество $\{x | h(x) \geq d\}$ и $\{x | g(x) \geq d\}$ отстоят друг от друга не более чем на ϵ . В этом случае нетрудно показать, что $g^-(\eta) = \inf_{\eta-\epsilon \leq x \leq \eta+\epsilon} g(x) \leq h(\eta) \leq \sup_{\eta-\epsilon \leq x \leq \eta+\epsilon} g(x) = g^+(\eta)$ для любых $h \in \mathcal{F}$.

Если бы функция правдоподобия h и параметр a были известны, степень реальности результатов наблюдений выражалась бы формулой

$\prod_i h[V_i - \Phi(X_i, a)]$ и оптимальная оценка получалась бы максимизацией этого выражения по a . Теперь, однако, ситуация иная, и мы должны рассчитывать на *наихудшее засорение*. Это приводит к следующему критерию оптимальной оценки:

$$\prod_i g^-[V_i - \Phi(X_i, a)] \Rightarrow \max. \quad (4.20)$$

Например, если $g(x) = \exp(-kx^2)$, то $g^-(\eta) = \inf_{\eta - \epsilon \leq x \leq \eta + \epsilon} \exp\{-kx^2\} = \exp\{-k(|\eta| + \epsilon)^2\}$, так что оптимальное a является решением задачи:

$$\sum_i [|V_i - \Phi(X_i, a)| + \epsilon]^2 \Rightarrow \min. \quad (4.21)$$

При $\epsilon \rightarrow 0$ этот метод переходит в метод наименьших квадратов, при $\epsilon \rightarrow \infty$ – в метод наименьших модулей, более устойчивый к грубым ошибкам. ■

4.3.2. Модель "больших" ошибок

Для "обычных", "малых" ошибок степень реальности убывает по мере возрастания величины ошибки. Таким образом, чем больше величина ошибки, то менее реальной она должна считаться. С "грубыми" ошибками положение должно быть обратным – чем больше величина ошибки, тем более реально, что она является "грубой". Поэтому для таких ошибок степень реальности (функция принадлежности или относительного правдоподобия) должна стремиться к 1 при неограниченном росте модуля величины ошибки и обращаться в нуль при нулевой величине ошибки. Этому отвечают, например, функции вида $p(\xi) = 2^{-D/|\xi|}$ или $p(\xi) = 2^{-(D/|\xi|)^2}$. Величина D здесь совпадает со средней (по модулю) ошибкой, т.е. такой, которой отвечает степень принадлежности или относительного правдоподобия, равная $1/2$. Представляется, что в каждом конкретном случае величина D может быть установлена исследователем хотя бы ориентировочно.

Пример 4.14. Оценивается параметр a в базовой модели в условиях, когда ошибки ξ , являются НВ и среди них имеются как "малые", так и "большие". Естественно считать, что функции принадлежности "малых" и "больших" ошибок четные и непрерывные, что позволяет обозначить их $h(|\xi|)$ и $H(|\xi|)$. Естественно также, что с ростом $|\xi|$ величина $h(|\xi|)$ убывает, а $H(|\xi|)$ – растет. Здесь функция принадлежности для всей совокупности наблюдений имеет вид:

$$\min \left\{ \min_{i \in \mathcal{K}} H\{Y_i - \Phi(X_i, a)\}, \min_{i \notin \mathcal{K}} h\{Y_i - \Phi(X_i, a)\} \right\}, \quad (4.22)$$

где \mathcal{K} – множество номеров наблюдений, которым отвечают "большие" ошибки.

В соответствии с МС-принципом оптимальная оценка отвечает

такому a и такому множеству \mathcal{H} , при которых выражение (4.22) максимально. Для выявления одного из условий оптимальности заметим, что уравнение $h(L) = H(L)$ имеет единственное положительное решение. Оно характеризует такую величину ошибки, которую с равной степенью принадлежности можно считать и "малой" и "грубой" – назовем ее "разделяющей". Пусть, например, функции принадлежности "малых" и "грубых" ошибок имеют вид: $h(\xi) = 2^{-(\xi/D)^2}$; $H(\xi) = 2^{-(R/\xi)^2}$, где D и R – соответствующие средние ошибки. Нетрудно проверить, что "разделяющая ошибка" здесь будет равна \sqrt{DR} – среднему геометрическому из средней "малой" и средней "грубой" ошибок. Легко проверить, что максимальное значение (4.22) будет достигаться, если в \mathcal{H} отнести все наблюдения, для которых $|Y_t - \Phi(X_t, a)| \geq L$. С другой стороны, пусть $|Y_t - \Phi(X_t, a)| < L$ для некоторого $t \in \mathcal{H}$. Если исключить это t из \mathcal{H} , то в силу неравенства $h\{Y_t - \Phi(X_t, a)\} > H\{Y_t - \Phi(X_t, a)\}$ критерий (4.22) увеличится или по крайней мере не уменьшится, что и требовалось доказать.

Мы видим, таким образом, что решение задачи "чрезмерно осторожно": оно требует минимизировать наибольшее отклонение, не превосходящее L , и максимизировать наименьшее отклонение, превосходящее L . Это неудивительно: степень принадлежности совокупности НВ определяется "наихудшей" из них, так что здесь МС-принцип не учитывает ни количества, ни величины других, "неэкстремальных" отклонений. ■

Пример 4.15. Пусть в той же задаче ошибки являются ПВ. Здесь критерий согласованности отражает относительное правдоподобие выборки и имеет вид:

$$\prod_{t \in \mathcal{H}} H\{Y_t - \Phi(X_t, a)\} \times \prod_{t \notin \mathcal{H}} h\{Y_t - \Phi(X_t, a)\} \Rightarrow \max.$$

Нетрудно показать, что этот критерий эквивалентен следующему:

$$\prod_t \max\{H\{Y_t - \Phi(X_t, a)\}, h\{Y_t - \Phi(X_t, a)\}\} \Rightarrow \max. \quad (4.23)$$

Повторяя предыдущие рассуждения, аналогично найдем, что множество \mathcal{H} содержит те и только те наблюдения, для которых $|Y_t - \Phi(X_t, a)| \geq L$. Однако здесь критерий учитывает все, а не только "экстремальные" (наиболее близкие к L) отклонения. В частном случае, когда "малые" и "большие" ошибки характеризуются функциями относительного правдоподобия $2^{-(\xi/R)^2}$ и $2^{-(S/\xi)^2}$, где R и S – средние значения соответствующих ошибок (см. п. 3.3), то критерий (4.23) принимает вид:

$$\sum_t \min \left\{ \frac{[Y_t - \Phi(X_t, a)]^2}{R^2}, \frac{S^2}{[Y_t - \Phi(X_t, a)]^2} \right\} \Rightarrow \min. \quad \blacksquare \quad (4.24)$$

4.3.3. Модель ошибок спецификации

В модели "засорения", собственно говоря, грубых ошибок нет – все они равноправны, однако закон их распределения не полностью согласован с исходной информацией. В модели "больших" ошибок предполагалось, что результат наблюдения является просто неточным (или даже *очень* неточным) измерением наблюдаемой характеристики объекта и потому, пусть плохо, отражает ее истинное значение. Поэтому здесь "грубая" ошибка отличается от "обычной" только своей величиной, но не содержанием. В то же время возможна и совершенно иная трактовка таких ошибок.

А именно, представим себе, что наблюдение производится путем выписывания данных из отчетности предприятия. Здесь не исключена ситуация, когда наблюдатель выпишет данные не из той клетки отчетной таблицы и выписанный показатель не будет иметь никакого отношения к требуемой характеристике объекта (например, вместо мощности объекта в одном случае будет выписан год ввода его в эксплуатацию, а в другом – численность персонала). Точно так же при отборе объектов наблюдения может случиться, что в число отобранных попадут те, которые на самом деле к ГС не относятся – и здесь соответствующие наблюдения вообще не имеют или имеют довольно слабое отношение к анализируемой ГС или к оцениваемой зависимости. Подобные ситуации в п. 1.3 трактовались как ошибки *спецификации*.

При наличии таких ошибок говорить о вероятностном распределении величины ошибок или о степени их реальности практически не имеет смысла – на первое место выходит проблема их устранения. Здесь полезно использовать указанную выше трактовку модели "засорения": с некоторой вероятностью наблюдаемая случайная величина имеет не заданное, а какое-то иное, засоряющее распределение. С другой стороны, критерием в модели "засорения" было наихудшее среднее значение смещение оцениваемого параметра, что явно предполагает повторяемость процедуры наблюдения и одновременно – повторяемость грубых ошибок с одним и тем же распределением. Это сразу же сужает круг задач, где соответствующие модели могут быть использованы. Между тем данный недостаток устраним – для этого следует поставить задачу иначе, сделав упор на разделении "грубых" и "обычных" ошибок.

Другими словами, надо рассмотреть совокупность наблюдаемых переменных как неоднородную, состоящую из двух групп: в одну входят показатели, которые действительно являются (содержащими ошибки) наблюдениями истинных характеристик объектов ГС, тогда как вторая состоит из "грубых ошибок", не имеющих к ГС никакого отношения. Искомая зависимость или ее параметры при этом определяются только по наблюдениям первой группы исходя из информации о степени реальности тех или иных ошибок наблюдения. Что же касается наблюдений второй группы (неверно специфицированных наблюдений), то они в расчете просто не участвуют. В то же время неизвестно,

какие именно наблюдения относятся к этой группе, и это восполняется информацией о степенях реальности ошибок спецификации.

Для *вероятностной* ситуации, когда распределение результатов наблюдения имеет плотность, построить такую модель нельзя (в одной формуле несовместимы вероятности одних ошибок и плотности вероятностей других). Однако положение меняется, если ошибки являются ПВ.

Пример 4.16. Пусть в базовой модели ошибки являются независимыми ПВ. Тогда модель "грубых" ошибок может устроена следующим образом. Каждое t -е наблюдение V_t характеризуется не только *величиной* ξ_t ошибки (отклонения V_t от истинного значения Y_t наблюдаемого показателя), но и ее *типом* τ_t . Значению $\tau_t = 0$ отвечают "обычные" ошибки, значению $\tau_t = 1$ – "грубые". При этом пары (ξ_t, τ_t) независимы, причем функции принадлежности для пар вида $(\xi, 0)$ и $(\xi, 1)$ различаются. Рассмотрим ситуацию, когда степень принадлежности для пары $(\xi, 0)$ равна $h(|\xi|)$, где h – убывающая функция, а все пары $(\xi, 1)$ при любых ξ имеют одну и ту же степень правдоподобия ϵ . Тогда функция относительного правдоподобия неизвестного параметра a принимает вид: $L = \epsilon^k \prod_{t \in \mathcal{K}} h(|\xi_t|)$, откуда вытекает следующий критерий оптимизации оценки a :

$$L = \epsilon^k \prod_{t \in \mathcal{K}} h(|V_t - \Phi(X_t, a)|) \Rightarrow \max.$$

Нетрудно убедиться, что для оптимального решения в множество \mathcal{K} будут отнесены только такие t , для которых $h(|V_t - \Phi(X_t, a)|) \leq \epsilon$. Поэтому полученный критерий оказывается эквивалентным следующему:

$$L = \prod_t \max\{h(|V_t - \Phi(X_t, a)|), \epsilon\} \Rightarrow \max. \blacksquare \quad (4.25)$$

4.4. Разнотипная неопределенность

При исследовании и оценке многофакторных зависимостей наблюдения разных переменных могут быть неопределенными величинами различных типов. Например, одни из них могут рассматриваться как СВ, другие – как НВ, третьи – как ПВ. Типичным примером является оценка параметров a и b в линейной зависимости $Y = aX + b$ в условиях, когда значения Y наблюдаются со случайными нормально распределенными ошибками с известной дисперсией, а наблюдаемые значения X – нечеткие с известной центрально-симметричной функцией принадлежности.

Другой вид "разнотипности" в рамках одной модели иллюстрируется следующим примером. Пусть ошибки наблюдения – СВ, имеющие нормальное распределение с нулевым средним и неточно известной

дисперсией S . Неопределенность величины S может носить различный характер, не обязательно вероятностный (так, может быть известен интервал, в котором лежит величина S). Конкретные подходы к решению этой проблемы пока неясны, однако имеет смысл привести примеры задач, где учесть разнотипность неопределенности удастся.

Пример 4.17. Пусть по T наблюдениям (Y_i, X_i) оцениваются параметры a и b в линейной зависимости $Y = aX + b$, причем переменная X наблюдается точно, а Y – со стохастической ошибкой ξ , имеющей нормальное распределение с нулевым средним и дисперсией S . Предположим также, что имеющаяся информация об ошибке наблюдения позволяет считать, что ее дисперсия S близка к известной величине D . Примем поэтому, что S является величиной, наделенной правдоподобием, а степень ее относительного правдоподобия принимает максимальное значение 1 в точке $S = D$. Будем считать, что она имеет вид

$\left(\frac{D}{S}\right)^q e^{q(1-D/S)}$, где q – некоторый параметр, определяющий ширину области достаточно правдоподобных значений S (при больших q функция правдоподобия резко уменьшается уже при малом отклонении S от D).

Если бы величина S была известна, то согласованность наблюдений (Y_i, X_i) с параметрами модели a и b характеризовалась "обычной"

функцией правдоподобия $\prod_{i=1}^T \frac{1}{\sqrt{2\pi S}} e^{-[Y_i - aX_i - b]^2 / 2S}$. Разделив эту величину на ее значение при наиболее правдоподобном значении $S = D$ и нулевых локальных отклонениях $Y_i - (aX_i + b)$, получим функцию относительного правдоподобия

$\prod_{i=1}^T \sqrt{\frac{D}{S}} e^{-[Y_i - aX_i - b]^2 / 2S}$. В такой ситуации

функция согласованности $\left(\frac{D}{S}\right)^q e^{q(1-D/S)} \times \prod_{i=1}^T \sqrt{\frac{D}{S}} e^{-[Y_i - aX_i - b]^2 / 2S}$ отразит

как правдоподобие оцененного значения дисперсии S , так и согласованность рассчитанных отклонений $Y_i - (aX_i + b)$ с принятой моделью.

Максимизируя этот критерий, получим, что оптимальные a и b обеспечивают минимум суммы квадратов отклонений

$\left(\sum_i [Y_i - aX_i - b]^2 \right)$, а оптимальная оценка дисперсии S дается формулой

$$S = \frac{2qD + \sum_i [Y_i - aX_i - b]^2}{2q + T}$$
. При малых q эта оценка близка к "обычной"

оценке максимального правдоподобия, при больших – к заданному, наиболее правдоподобному значению D . ■

Пример 4.18. Оценивается параметр a зависимости $Y = \Phi(X, a)$,

связывающей истинные значения показателей некоторых объектов, причем истинные значения X_i измеряются точно, а Y_i – с аддитивной ошибкой ξ_i , наделенной правдоподобием. Основная часть ошибок – "малые", они характеризуются функцией относительного правдоподобия $e^{-b\xi^2}$. Однако некоторые ошибки могут быть "грубыми", для них функция относительного правдоподобия $h(\xi)$ неизвестна. Примем, что она зависит от $|\xi|$ и растет при увеличении $|\xi|$. Количество грубых ошибок – тоже ПВ. Если бы такие ошибки возникали случайно, с вероятностью ε в каждом наблюдении, то вероятность появления k ошибок в T наблюдения составляла бы $C_T^k \varepsilon^k (1-\varepsilon)^{T-k}$. Примем поэтому, что количество грубых ошибок характеризуется степенью правдоподобия, пропорциональной этому выражению, а размеры грубых ошибок не зависят от их количества. Тогда, обозначив через \mathcal{K} множество номеров наблюдений, отвечающих грубым ошибкам, а через $k = |\mathcal{K}|$ – количество таких наблюдений, получим следующее выражение для критерия согласованности:

$$L = C_T^k \varepsilon^k (1-\varepsilon)^{T-k} \prod_{i \in \mathcal{K}} e^{-b[Y_i - \Phi(X_i, a)]^2} \prod_{i \notin \mathcal{K}} h(|Y_i - \Phi(X_i, a)|) \Rightarrow \max. \quad (4.26)$$

Выясним свойства оптимального решения, зафиксировав вначале количество грубых ошибок. Предположим, что t -е наблюдение отвечает грубой ошибке, а s -е – "малой" ошибке. Тогда значение L не должно увеличиться, если в множестве \mathcal{K} заменить t на s . Легко видеть, что это будет только, если

$$\begin{aligned} & b[Y_t - \Phi(X_t, a)]^2 + \ln h(|Y_t - \Phi(X_t, a)|) \geq \\ & \geq b[Y_s - \Phi(X_s, a)]^2 + \ln h(|Y_s - \Phi(X_s, a)|). \end{aligned}$$

Но при положительных z функция $bz^2 + \ln h(z)$ – возрастающая, поэтому данное неравенство возможно только, если $u_t = |Y_t - \Phi(X_t, a)| \geq |Y_s - \Phi(X_s, a)| = u_s$. Таким образом, для всех $t \in \mathcal{K}$ локальные отклонения u_t не меньше, чем для всех остальных i . Это означает, что грубыми ошибками следует считать здесь те k наблюдений, которым отвечают наибольшие по модулю локальные отклонения. Обозначим поэтому k -е по порядку убывания из локальных отклонений через $u_{(k)}$.

Функция правдоподобия для грубых ошибок нам неизвестна, но ее оптимальная оценка, отвечающая максимальному значению критерия (4.26), будет очень простой: $h(\xi) = 1$ для $|\xi| \geq u_{(k)}$. Тогда критерий (4.26) примет вид:

$$L = C_T^k \varepsilon^k (1-\varepsilon)^{T-k} \prod_{i \in \mathcal{K}} e^{-b[Y_i - \Phi(X_i, a)]^2} \Rightarrow \max. \quad (4.27)$$

Отсюда следует, что оптимальная оценка параметра a может быть получена методом наименьших квадратов исходя из наблюдений, не входящих в множество \mathcal{H} . Таким образом: 1) множество \mathcal{H} отвечает k наибольшим локальным отклонениям $u_i = |Y_i - \Phi(X_i, a)|$; 2) оптимальное значение a обеспечивает минимум суммы $S(\mathcal{H}) = \sum_{i \in \mathcal{H}} [Y_i - \Phi(X_i, a)]^2$;

3) оптимальное количество грубых ошибок k отвечает максимуму выражения $C_T^k \varepsilon^k (1 - \varepsilon)^{T-k} e^{-bS(\mathcal{H})}$. Эти свойства могут быть положены в основу итеративной процедуры оптимального оценивания. ■

В примере 2.8 мы рассмотрели задачу оценки параметра в условиях неоднородной ГС и стохастических ошибок. Аналогичный метод применим и тогда, когда ошибки являются ПВ или НВ.

Пример 4.19. Имеются данные о двух типах объектов. Наблюдаемые характеристики Y_i и X_i объектов первого типа связаны зависимостью $Y_i = \alpha X_i + b + \xi_i$ ($i = 1, \dots, T$), для объектов второго типа имеет место аналогичная зависимость $V_k = \alpha U_k + d + \eta_k$ ($k = 1, \dots, K$). Ошибки ξ_i и η_k являются независимыми ПВ с известными функциями правдоподобия, соответственно $e^{-\xi^2/R}$ и $e^{-\eta^2/S}$, где R и S – средние значения квадратов ошибок, умноженные на $\ln 2$. Требуется оценить общий для обеих зависимостей параметра α . В этой ситуации критерий оптимальной оценки имеет вид:

$$L = e^{-R^{-1} \sum_i (\alpha X_i - Y_i + b)^2 - S^{-1} \sum_k (\alpha U_k - V_k + d)^2} \Rightarrow \max.$$

Без информации о R и S эту задачу решить нельзя, однако знать их точные значения не обязательно. Пусть, например, R и S неизвестны, но их можно рассматривать как ПВ *априорно* известными наиболее правдоподобными значениями G и H . Примем, что соответствующие

функции правдоподобия имеют вид $\left(\frac{G}{R}\right)^q e^{q(1-G/R)}$ и $\left(\frac{H}{S}\right)^q e^{q(1-H/S)}$.

При этом оптимальное α определяется из условия:

$$L = \left(\frac{GH}{RS}\right)^q \times \exp \left[q \left(2 - \frac{G}{R} - \frac{H}{S} \right) - R^{-1} \sum_i (\alpha X_i - Y_i + b)^2 - S^{-1} \sum_k (\alpha U_k - V_k + d)^2 \right] \Rightarrow \max. \quad \blacksquare$$

4.5. Устранение влияния отдельных факторов

Как говорилось в п. 1.5.2, нередко возникает потребность в установлении зависимости объясняемой переменной Y от объясняющей переменной X в условиях, когда заведомо известно, что на Y влияют и

иные факторы. Примеры оценки такого рода *неполных* зависимостей приводятся ниже.

Пример 4.20. Рассмотрим задачу выявления сезонных колебаний в динамике какого-либо "объемного" показателя X (например, объема производства на данном предприятии). Такая задача представляет интерес не только потому, что влияние сезонных факторов интересно само по себе. Нередко установление многих эконометрических зависимостей затруднено малым объемом исходной информации, и использование помесечных или поквартальных данных позволило бы повысить точность соответствующих оценок, если бы только из этих данных можно было бы исключить влияние сезонных факторов. Покажем, как можно это сделать.

Пусть известны значения Y_t анализируемого показателям в t -м квартале достаточно большого отчетного периода ($t = 1, \dots, T$), начинающегося с I квартала некоторого календарного года. Предполагается, что в значениях Y могут быть выделены сезонная и "трендовая" ("очищенная от влияния сезонности") составляющие. Последняя характеризуется динамикой (неизвестных) "трендовых" значений Z_t , причем эта динамика должна быть "нормальной", влияние же сезонных факторов приводит к повышению или понижению фактических значений Y по отношению к "трендовым" в определенное число раз, зависящее только от номера квартала в соответствующем году. Иными словами, имеет место разложение:

$$Y_t = Z_t \times J_t, \quad (4.28)$$

где J_t – коэффициент сезонного изменения Y (коэффициент сезонности), относящийся к t -му кварталу [44]. Эти коэффициенты через четыре квартала повторяются, так что их динамика периодическая:

$$\begin{aligned} J_1 &= J_5 = J_9 = \dots; J_2 = J_6 = J_{10} = \dots; \\ J_3 &= J_7 = J_{11} = \dots; J_4 = J_8 = J_{12} = \dots \end{aligned} \quad (4.29)$$

Однако только этого для оценки зависимости недостаточно. Предположим, что в течение всех лет расчетного периода динамика Y была следующей: в течение первых трех кварталов $Y = 10$, в IV квартале $Y = 20$. Спрашивается, что мы имеем: сезонный рост в четвертом квартале или сезонное снижение в первых трех или то и другое вместе?

Чтобы ответить на этот вопрос, надо договориться о том, какому кварталу будет отвечать индекс сезонности, равный единице. Пусть, для определенности, таким кварталом будет первый. Это даст нам дополнительное соотношение:

$$J_1 = 1. \quad (4.30)$$

Формализуем теперь понятие "тренда". Выявляя влияние сезонных факторов, мы одновременно хотим и "очистить" от этого влияния динамику исследуемого показателя, т.е. сделать динамику трендовых значений Z_t возможно более "гладкой", "плавной". Тем самым крите-

рием оптимального разложения должен выступать некоторый показатель "плавности" указанной динамики. В непрерывном случае здесь можно было бы использовать описанный в п. 2.2 интегральный критерий Q_2 . Поскольку в нашей задаче время меняется дискретно, воспользуемся дискретным аналогом указанного критерия. Это приводит к задаче нахождения неизвестных J_t и Z_t из условий (4.28)–(4.30) и

$$\sum_{t=2}^{T-1} (Z_{t+1} - 2Z_t + Z_{t-1})^2 \Rightarrow \min. \quad (4.31)$$

Легко видеть, что это эквивалентно нахождению коэффициентов J_t из условия

$$\sum_{t=2}^{T-1} \left(\frac{Y_{t+1}}{J_{t+1}} - 2 \frac{Y_t}{J_t} + \frac{Y_{t-1}}{J_{t-1}} \right)^2 \Rightarrow \min \quad (4.32)$$

при ограничениях (4.29)–(4.30). Нетрудно убедиться, что оптимальные коэффициенты J_t (точнее, обратные к ним величины) находятся теперь из решения системы трех линейных уравнений с тремя неизвестными, после чего "очищенные" от влияния сезонности значения показателя (Z_t) находятся из формулы (4.28). ■

Пример 4.21. При построении производственной функции, отражающей зависимость *максимально возможного* объема производимого продукта от имеющегося количества ресурсов, требуется скорректировать ВВП с учетом "недоиспользования производственной мощности народного хозяйства". Обычно в этих целях используется метод Уортонской школы "тренд через пики" [66]. Он состоит в проведении линии тренда по пиковым точкам временного ряда объемов ВВП, причем линия тренда отвечает *потенциальному* ВВП, а уровень использования мощностей – отношению фактического ВВП к "трендовому". Такому методу присущи два недостатка (см. также [4]): 1) в "пиковых" точках ВВП предполагается отвечающим полной загрузке мощностей; 2) предполагается, что линейная интерполяция между пиковыми точками является хорошим приближением темпов роста мощностей. Оказывается, есть возможность, хотя бы частично, устранить второй недостаток (устранить первый, по-видимому, нельзя даже, если рассматривать объемы производства в отдельных отраслях экономики).

Действительно, пусть D_t – фактический ВВП в году t , а Y_t – потенциальный ВВП (мощность народного хозяйства по производству ВВП) в этом году. Тогда величина $Z_t = D_t/Y_t$ отражает уровень загрузки мощностей в этом году. Рассмотрим логарифмы указанных величин: $d_t = \ln(D_t)$, $y_t = \ln(Y_t)$, $z_t = -\ln(Z_t) = y_t - d_t$. Будем рассматривать величины Z_t и зависимость $y_t = f(t)$ как наделенные правдоподобием.

С одной стороны, представляется весьма правдоподобным, что объемы потенциального ВВП с течением времени растут, причем достаточно плавно, так что функция $f(t)$ возрастающая и гладкая в смысле п. 2.4. Тогда, как и в рассмотренной выше ситуации, функцию

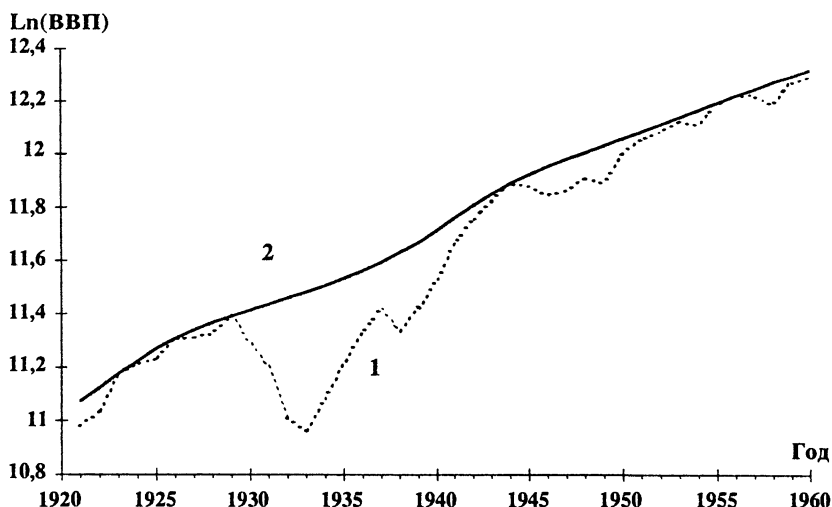


Рис. 3. Оценка потенциального ВВП
1 — факт; 2 — расчет

правдоподобия для нее можно задать через дискретный аналог функционала гладкости (2.5): $\exp\left\{-\lambda \sum_{t=2}^{T-1} \left[1 - \frac{y_{t+1} - y_t}{y_{t+1} - y_t}\right]^2\right\}$.

С другой стороны, естественно ожидать, что малые уровни загрузки мощностей значительно менее правдоподобны, чем близкие к единице. На этом основании примем, что Z_t имеют на отрезке $[0, 1]$ функцию правдоподобия $p(Z) = Z^m$. При этом величины z_t будут иметь на $[0, \infty)$ функцию правдоподобия $p(z) = e^{-mz}$. Тогда функция правдоподобия результатов наблюдений станет равной

$$\exp\left\{-\lambda \sum_{t=2}^{T-1} \left[1 - \frac{y_{t+1} - y_t}{y_{t+1} - y_t}\right]^2 - m \sum_{t=1}^T z_t\right\}.$$

Введя параметр $\mu = \lambda/m$ и учитывая, что $z_t = y_t - d_t \geq 0$, задачу максимизации этого критерия сведем к следующей:

$$\mu \sum_{t=2}^{T-1} \left[\frac{y_{t+1} - y_t}{y_{t+1} - y_t} - 1 \right]^2 + \sum_{t=1}^T (d_t - y_t) \Rightarrow \max, \quad y_t \leq d_t \quad \forall t. \quad (4.33)$$

На рис. 3 приведена динамика фактического и потенциального ВВП по частному несельскохозяйственному сектору США за 1921–1960 гг., полученная из (4.33) при $\mu = 2$ на основе данных [4]. Рисунок показывает, что общая идея проведения "тренда через пики" при использовании данного метода сохраняется. Расчеты показывают, кстати, что

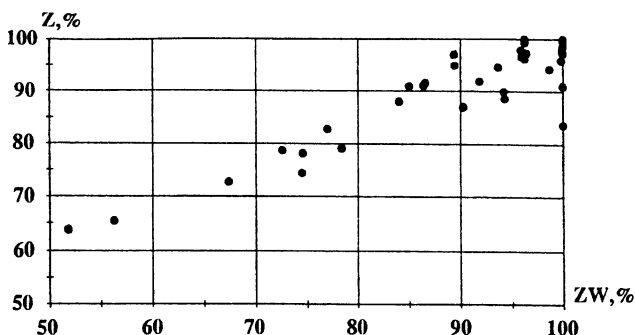


Рис. 4. Сопоставление уровней загрузки мощностей

динамика потенциального ВВП оказывается довольно устойчивой к изменению параметра μ . Интересно сравнить найденные коэффициенты загрузки (Z) с рассчитанными методом Уортонской школы (ZW), значения которых также приведены в [4]. Результаты такого сопоставления представлены на рис. 4. Из рисунков видно, что предлагаемый метод по сравнению с методом Уортонской школы дает в среднем более высокие значения уровня загрузки мощностей, при этом динамика потенциального ВВП, исчисленная вторым способом, представляется более удовлетворительной. Отметим в этом связи, что использование вспомогательного параметра μ , хотя и вносит в решение определенный элемент субъективизма, позволяет учесть при решении особенности задачи, не формализованные в ее постановке и "ощущаемые" исследователем на интуитивном уровне. ■

Подобные подходы применимы, по нашему мнению, и в других задачах "очистки" наблюдаемой переменной от влияния тех или иных "вспомогательных" факторов.

ЗАКЛЮЧЕНИЕ

Проблемы построения моделей зависимостей между экономическими показателями, так же как и иных экономико-математических моделей, в настоящее время далеки от окончательного решения. Они, к сожалению, в последние десятилетия отодвинулись на периферию интересов многих специалистов (в числе немногих исключений следует упомянуть вероятностную ветвь прикладной статистики). В известных публикациях основное внимание уделяется не методологии, а построению и исследованию конкретных (специфицированных или нет) моделей.

Между тем отсутствие единой и обоснованной методологии и методики моделирования создает ситуацию, когда различные модели одного и того же объекта с трудом поддаются сравнению, проверке на адекватность, не допускают объединения в едином комплексе т.д. В результате рекомендации, выдаваемые на основе модельных расчетов, часто носят весьма субъективный характер (достаточно вспомнить недавнюю дискуссию по поводу эконометрических моделей зависимости уровня экономической динамики страны от размеров государственного вмешательства в экономику: одни публикации подтверждали необходимость редукции такого вмешательства, другие – целесообразность его расширения).

Не меньшую опасность, чем *субъективизм* в построении и интерпретации моделей, представляет собой другое явление в сфере математического моделирования, которое можно назвать *конвенционализмом*. Речь идет о том, что многие традиционные методы интерпретации и обработки данных, разработанные для вполне определенных ситуаций моделирования, включающих специфическую предметную область и цели исследования по молчаливо принимаемому соглашению распространяются на совершенно другие ситуации. Конвенционализм в моделировании, направленный в целом на преодоление субъективизма, тем не менее не снимает его полностью и является в определенном смысле не меньшим злом, поскольку затрудняет возможность привлечения внимания исследователей к проблемам повышения объективности построения и применения экономико-математических моделей.

Мы не ставили задачу заполнить весь имеющийся в данной проблематике пробел. За пределами исследования остались вопросы выбора систем факторных показателей, методика и границы экономической интерпретации тех или иных характеристик построенных зависимостей полностью и другие вопросы. Словом, вопросов осталось больше, чем получено ответов. Вместе с тем мы надеемся, что данная работа будет способствовать пробуждению интереса у специалистов по эконометрике, статистике, другим экономико-математическим дисциплинам к фундаментальным проблемам методологии моделирования, консолидации усилий исследователей в актуальном для развития экономико-математических методов научном направлении.

ЛИТЕРАТУРА

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справ. изд. М.: Финансы и статистика, 1985.
2. Кендалл М.Дж., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973.
3. Клейнер Г.Б. Детерминированный анализ систем показателей // Экономика и мат. методы. 1981. Т. XVII. Вып. 6.
4. Браун М. Теория и измерение технического прогресса. М.: Статистика, 1971.
5. Тинтнер Г. Введение в эконометрию. М.: Статистика, 1965.
6. Клейнер Г.Б. Производственные функции: теория, методы, применение. М.: Финансы и статистика, 1986.
7. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: Юнити, 1998.
8. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. М.: Финансы и статистика, 1986.
9. Пфанцagl И. Теория измерений. М.: Мир, 1976.
10. Чупров А.А. Очерки по теории статистики. М.: Госстатиздат, 1959.
11. Раяцкас Р.Л., Плакунов М.К. Количественный анализ в экономике. М.: Наука, 1987.
12. Шклярский Д.О., Ченцов Н.Н., Яглом И.М. Геометрические оценки и задачи комбинаторной геометрии. М.: Наука, 1974.
13. Теория ошибок // Весь компьютерный мир, 1996. № 1.
14. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1968.
15. Смоляк С.А. Планирование многономенклатурного производства. М.: Экономика, 1977.
16. Смоляк С.А. Как обосновать вид зависимости между экономическими показателями // Наука и технология в России. 1994. № 1(3). С. 11–12.
17. Боярский А.Я. и др. Общая теория статистики / Учебник. М.: Финансы и статистика, 1985.
18. Методы анализа данных: Подход, основанный на методе динамических сгущений (под ред. Дидэ Э.). М.: Финансы и статистика, 1985.
19. Benzekri J.P. et al. L'analyse des donnees. Dunod, 1973. Tome 1; Dunod, 1982. Tome II.
20. Брейли Р., Майерс С. Принципы корпоративных финансов. М.: Олимп-бизнес, 1997.
21. Бахвалов Н.С. Численные методы. М.: Наука, 1973.
22. Kolmogorov A.N. Über die beste Annäherung von Funktionen einer gegebenen Funktionenklasse // Ann. of Math. 1936. V. 37.
23. Никольский С.М. Квадратурные формулы. М.: Физматгиз, 1958 (издание второе с добавлением Н.П. Корнейчука. М.: Наука, 1974).
24. Тихомиров В.М. Поперечники множеств в функциональном пространстве и теория наилучших приближений // Успехи мат. наук. 1960. Т. 14. № 3.

25. Тихомиров В.М. Наилучшие методы приближения и интерполирования дифференцируемых функций в пространстве $C(-1,1)$ // Мат. сб. 1969. Т. 80 (122). № 2(10).
26. Смоляк С.А. Интерполяционные и квадратурные формулы на классах W_s^α и E_s^α // Доклады АН СССР. 1960. Т. 131. № 5.
27. Смоляк С.А. Об оптимальном восстановлении функций и функционалов от них. Дис. канд. экон. наук. М., 1966.
28. Тихомиров В.М. Некоторые вопросы теории приближений. М.: Изд-во МГУ, 1976.
29. Клейнер Г.Б. Область определения производственной функции // Экономика и мат. методы. 1978. Т. XIV. Вып. 5.
30. Uzawa H. Production Functions with Constant Elasticities of Substitution // The Review of Economic Studies. V. 29 (October, 1962). P. 91–99.
31. Mukerji V. Generalized SMAC Functions with Constant Ratios of Elasticities of Substitution // The Review of Economic Studies. V. 30 (October, 1963). P. 233–261.
32. Клейнер Г.Б., Сирота Б.Н. О производственных функциях с постоянными и переменными эластичностями замены факторов // Экономика и мат. методы. 1975. Т. XI. Вып. 3.
33. Демиденко Е.З. Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981.
34. Holladay J.C. Smoothest curve approximation // Math. Tables Aids Comput. 1957. V. 11.
35. Смоляк С.А. Сплаины и их применение // Экономика и мат. методы. 1971. Т. VII. Вып. 3.
36. Алберг Дж., Нильсен Э., Уолш Дж. Теория сплайнов и ее приложения. М.: Мир, 1972.
37. Рябушкин Т.В. и др. Общая теория статистики // Учебник. М.: Финансы и статистика, 1981.
38. Fisher R.A. On the mathematical foundations of theoretical statistics. // Phil. Trans., 1921. Sec. A. Vol. 222.
39. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983.
40. Lindley D.V. Regression lines and the linear functional relationship // Suppl. J. Roy. Statist. 1947. Vol. 9.
41. Kiefer J., Wolfowitz J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters // Ann. Math. Statist. 1956. V. 27.
42. Хьюбер П. Робастность в статистике. М.: Мир, 1984.
43. Феллер В. Введение в теорию вероятностей и ее приложения. М.: Мир, 1967. Т. 2.
44. Смоляк С.А., Титаренко Б.П. Устойчивые методы оценивания (статистическая обработка неоднородных совокупностей). М.: Статистика, 1980.
45. Neyman J., Pearson E.S. On the use and interpretation of certain test criteria for the purposes of statistical inference // Biometrika. 1928. V. 20A.
46. Клейнер Г.Б., Николаева Н.Л. Оценка параметров имитационных экономико-статистических моделей с учетом априорной качественной информации // Экономика и мат. методы. 1986. Т. XXII. Вып. 4.

47. Курно О. Основы теории шансов и вероятностей. М.: Наука, 1970.
48. Карпенко Б.И. Случайность и необходимость и их взаимосвязи в статистических исследованиях // Проблемы теории статистики. М.: Наука, 1978.
49. Savage L.J. The Foundations of Statistics. N.Y.: Wiley, 1954.
50. Фишберн П. Теория полезности для принятия решений. М.: Наука, 1978.
51. Полтерович В.М. Кризис экономической теории // Экономическая наука современной России. 1998, № 1.
52. Fama E.F., French K.R. CAPM is wanted, dead or alive // J. of Finance. 1995, October.
53. Литтлвуд Дж. Математическая смесь. М.: Физматгиз, 1962.
54. Zadeh L.A. Fuzzy Sets // Information and Control. 1965. Vol. 8.
55. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь. 1982.
56. Нечеткие множества в моделях управления и искусственного интеллекта (под ред. Поспелова Д.А.). М.: Наука, 1986.
57. Обработка нечеткой информации в системах принятия решений. М.: Радио и связь. 1989.
58. Смоляк С.А. О правилах сравнения нечетких альтернатив // Экономика и мат. методы. 1993. Т. 29. Вып. 4.
59. Левнер Е.В., Птускин А.С., Фридман А.А. Размытые множества и их применение. М.: ЦЭМИ РАН, 1998.
60. Смоляк С.А. О сравнении альтернатив, параметры которых характеризуются функциями правдоподобия // Экономика и мат. методы. 1996. Т. 32. Вып. 1.
61. Пойа Д. Математика и правдоподобные рассуждения. М.: Наука, 1975.
62. Алексеев А.В. и др. Интерпретация значений функции принадлежности и операции над нечеткими множествами // Прикладные задачи анализа решений в организационно-технических системах. Рига: Рижский политехнический институт, 1983.
63. Смоляк С.А. О правилах сравнения альтернатив с неопределенными затратами и результатами // Вероятностные модели математической экономики. М.: ЦЭМИ, 1990.
64. Смоляк С.А. Оценка эффективности проектов в условиях интервально-вероятностной неопределенности // Экономика и мат. методы, 1998. Т. 34. Вып. 3.
65. Pearson K. On lines and planes closest fit to systems of points in space // Phil. Mag. 1901. Vol. 2. S. 6.
66. Index of Percentage Utilization of Industrial Capacity for the United States // Econometric Research Unit. University of Pennsylvania, 1960.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1. ОСНОВНЫЕ ПОНЯТИЯ	5
1.1. Объекты и их характеристики	5
1.2. Наблюдения, измерения, оценки	7
1.3. Ошибки наблюдений	11
1.4. Зависимости между характеристиками объектов	15
1.5. Различные типы статистических зависимостей	17
2. ПРИНЦИПЫ ОПТИМАЛЬНОЙ ОЦЕНКИ ЗАВИСИМОСТЕЙ	23
2.1. Принцип минимального расстояния	23
2.2. Принцип минимальных потерь	28
2.3. Принцип минимума интегрального отклонения	29
2.4. Принцип максимальной гладкости в задаче интерполяции	34
2.5. Принцип максимального правдоподобия	36
2.6. Два правдоподобия	42
2.7. Основные недостатки метода максимального правдоподобия ...	47
2.8. Сочетание разных принципов	50
2.9. Принцип максимальной согласованности	51
2.10. Об учете целевой информации при оценивании зависимостей.	55
3. ХАРАКТЕРИЗАЦИЯ НЕОПРЕДЕЛЕННОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ	57
3.1. Систематические ошибки	57
3.2. "Случайные" результаты наблюдений – вероятностный подход.	58
3.3. Альтернативные описания "случайных" величин	67
3.4. Интервальные неопределенности	73
4. ОЦЕНКА ЗАВИСИМОСТЕЙ ПРИ РАЗНЫХ ТИПАХ НЕОПРЕ- ДЕЛЕННОСТИ	75
4.1. Оценка параметрических зависимостей	75
4.2. Оценка непараметрических зависимостей	80
4.3. Выявление грубых ошибок и устранение их влияния	86
4.4. Разнотипная неопределенность	92
4.5. Устранение влияния отдельных факторов	95
ЗАКЛЮЧЕНИЕ	99
ЛИТЕРАТУРА	101