

ЭКОНОМИКО –
СТАТИСТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
В ПРОМЫШЛЕННОСТИ



ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ

АКАДЕМИЯ НАУК СССР
СИБИРСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ ЭКОНОМИКИ И ОРГАНИЗАЦИИ
ПРОМЫШЛЕННОГО ПРОИЗВОДСТВА

ЭКОНОМИКО-СТАТИСТИЧЕСКОЕ
МОДЕЛИРОВАНИЕ
В ПРОМЫШЛЕННОСТИ
(методологические и методические
вопросы)



ИЗДАТЕЛЬСТВО «НАУКА»
СИБИРСКОЕ ОТДЕЛЕНИЕ
Новосибирск · 1977

Монография посвящена комплексу методологических и методических вопросов статистического моделирования экономических показателей работы промышленных объектов. В первом, методологическом разделе излагаются методики построения различных типов экономико-статистических моделей: дискретных, дискретно-непрерывных статических и динамических; исследуются вопросы их оптимизации. Во втором разделе рассматриваются проблемы применения статистических моделей в анализе, нормировании и планировании промышленного производства, предлагается комплекс специальных прикладных методик решения типичных экономических задач.

Книга рассчитана на широкий круг специалистов, интересующихся применением статистических методов в анализе и планировании работы промышленных предприятий: экономистов, инженеров, сотрудников научно-исследовательских учреждений и плановых органов.

Редколлегия: *М. Л. Лукацкая, Н. В. Овчинникова, Б. Б. Розин* (ответственный редактор).

ВВЕДЕНИЕ

Современный этап развития народного хозяйства предъявляет новые требования к методам управления экономикой и среди них — к экономико-математическому моделированию. Как подчеркивалось в Отчетном докладе ЦК КПСС XXIV съезду партии, «наука серьезно обогатила теоретический арсенал планирования, разработав методы экономико-математического моделирования, системного анализа и другие. Необходимо шире использовать эти методы...»¹.

С развитием производства задача оценки и прогнозирования экономических показателей становится неразрешимой без коренного совершенствования методологии и методики их изучения. Применение методов статистического моделирования — один из реальных и эффективных путей совершенствования информационного обеспечения работ по планированию и управлению промышленным производством.

Общий научный задел по использованию методов математической статистики в экономических исследованиях создан трудами экономистов-статистиков: В. Е. Адамова, А. Я. Боярского, Г. И. Бакланова, А. Л. Вайнштейна, И. Г. Венецкого, С. М. Вишнева, Я. П. Герчука, А. М. Длинна, Н. К. Дружинина, Г. С. Кильдишева, Я. И. Лукомского, П. П. Маслова, В. С. Немчинова, И. С. Пасхавера, П. М. Рабиновича, А. И. Ротштейна, Т. В. Рябушкина, В. И. Сисьякова, Г. С. Струмилина, И. П. Сулова, Н. С. Чет-

¹ Материалы XXIV съезда КПСС. М., Госполитиздат, 1971, с. 67.

верикова, А. А. Чупрова, Б. С. Ястремского и других.

Однако, несмотря на наличие благоприятных объективных условий, методы статистического моделирования в оценке и прогнозировании показателей работы промышленных объектов до настоящего времени применяются недостаточно. Одной из причин этого представляется нам неразработанность некоторых общих методологических проблем экономико-статистического моделирования (например, проблемы учета типологической структуры совокупности или динамических аспектов процесса производства), а также отсутствие доступных и апробированных на практике общих и специальных методик решения типичных задач анализа, нормирования и планирования промышленного производства на основе экономико-статистических моделей.

В настоящей работе предпринята попытка создания таких комплексных методических рекомендаций, в которых были бы отражены как общеметодологические вопросы экономико-статистического моделирования промышленного производства, так и методические подходы к решению наиболее типичных экономических задач.

Основным объектом моделирования выбрано промышленное предприятие. Дело в том, что в общегосударственной системе экономических показателей особую роль играют показатели работы именно промышленных предприятий, представляющих один из важнейших уровней организации народного хозяйства. Предприятие не только производит материальные ценности, но и выступает источником первичной информации об условиях и результатах материального производства. На основе этой информации строятся агрегированные показатели более высоких уровней управления народным хозяйством.

Построение моделей формирования экономических показателей работы предприятий — необходимый составной элемент общей системы моделей территориально-производственного планирования. В конечном итоге любая система — отраслевая, районная, территориально-производственный комплекс — складывается из совокупности тех или иных предприятий — промышленных объектов. Большинство мероприятий по совершенствованию техники, технологии и управления производством осуществляется на предприятиях и тем или иным способом проявляется в их экономических показателях.

Для построения моделей функционирования предприятий могут использоваться различные подходы: методы сетевого и матричного планирования, имитационный и статистические подходы и т. п. В настоящей работе излагаются методические вопросы построения и применения моделей объектов последним из перечисленных способов — на основе экономико-статистического моделирования.

Элементы статистического подхода в той или иной мере присущи любому способу построения моделей промышленных предприятий (вернее, моделей, имитирующих механизм формирования технико-экономических показателей предприятий). В частности, при построении сетевых моделей статистические методы используются для оценки длительности работ и величины затрат; применяемые в матричном планировании нормативы трудовых и материальных затрат, как правило, имеют статистическую природу; при имитации производственного процесса на ЭВМ статистические методы используются не только в разработке нормативов, но и сама процедура решения задачи носит статистический характер (метод Монте-Карло).

В отличие от этих «опосредованных» способов использования статистического аппарата мы ориентируемся на прямое математико-статистическое моделирование механизма формирования экономических показателей работы предприятий.

При таком подходе для оценки значений экономических показателей используются экономико-статистические модели (ЭСМ) функционирования промышленных объектов (предприятия, цехи, крупные технологические агрегаты), описывающие зависимость результатов их деятельности (выпуск продукции, себестоимость, трудоемкость, фондоемкость и т. п.) от производственных факторов (объем и структура основных фондов, качество ресурсов, природные условия и т. д.).

С развитием методов математической статистики моделирование экономических явлений получило надлежащую статистическую основу. Здесь на первый план выдвинута идея статистических связей и взаимозависимостей. Статистический подход дает возможность учесть неопределенность информации, обойтись без специальных экспериментов, которые почти нельзя провести на уровне предприятия; он может эффективно использоваться на разных этапах экономического исследования.

Методы математической статистики позволяют изучать закономерности промышленного производства в обычных условиях. Данные производственно-технической документации рассматриваются как ранее поставленные массовые опыты. Поэтому во многих случаях отпадает необходимость проводить длительные и дорогостоящие эксперименты, к тому же нередко невозможные.

Статистические методы применимы в изучении опосредованных связей показателей и факторов, т. е. таких связей, когда имеется большая цепь причин и следствий между факторами и результатами и осуществить «прямой» расчет невозможно.

Актуальность проблемы экономико-статистического моделирования работы предприятий особенно возросла на современном этапе развития промышленного производства, когда значительно увеличились его масштабы (в народном хозяйстве СССР сейчас функционирует около 50 тыс. промышленных предприятий), повысилась техническая оснащенность, быстрыми и все более ускоряющимися темпами развивается технический прогресс, лавинообразно нарастает объем информации.

Обработка методологии построения адекватных ЭСМ объектов позволяет решить много общих проблем экономико-статистического моделирования и прогнозирования. Наиболее сложно с точки зрения статистической методологии моделировать и прогнозировать показатели объектов низкого уровня, развитие которых не носит плавного характера и включает значительный элемент неопределенности. Чем выше уровень моделируемого объекта, чем больше единиц и периодов времени осредняется, тем более плавная тенденция развития имеет место, тем лучше она поддается статистическому моделированию и прогнозированию. Поэтому с освоением методики моделирования и прогнозирования работы отдельных промышленных объектов, по существу, создается большой общеметодический задел для экономико-статистического моделирования вообще.

Излагаемые методические рекомендации разработаны на основе изучения специальной экономической, статистической и математической литературы и обобщения опыта научных исследований, проводимых в Лаборатории экономико-математических исследований Новосибирского государственного университета (1963—1966 гг.) и в Институ-

те экономики и организации промышленного производства СО АН СССР (с 1966 г.)². В исследованиях нашли отражение проблемы как внутривзаводского экономико-статистического анализа (объект исследования — отдельные технологические агрегаты и участки), так и межзаводского (объект исследования — предприятия в целом). За указанный период решено около 25 крупных прикладных задач по металлургической, лесозаготовительной, угольной, химической, цементной и другим отраслям промышленности.

Для внутривзаводских исследований характерны, во-первых, использование информации в виде динамических рядов о работе отдельных агрегатов или их совокупностей в пределах одного участка, цеха; во-вторых, узкая область приложимости получаемых ЭСМ — только к исследуемому объекту. Для подобных задач, как правило, не возникают проблемы изучения однородности совокупности, используются регрессионные модели непрерывного типа и наибольшая сложность связана с динамическим характером информации и технологическими особенностями объекта.

К этой группе относится цикл исследований на Кузнецком и Магнитогорском металлургических комбинатах, Карагандинском заводе синтетического каучука (СК), проведенных в 1965—1967 гг. На металлургических комбинатах изучалась работа мартеновских печей для целей предсказания длительности отдельной плавки и основных ее технологических периодов на базе пассивного эксперимента. Для Карагандинского завода СК решалась задача моделирования технико-экономических показателей работы цехов по производству карбида кальция, ацетальдегида и этилацетата с последующей оптимизацией рабочих режимов на основе получаемого экономико-статистического описания.

В ходе этих исследований решены методические вопросы построения моделей работы крупного технологического агрегата, основанных на использовании регрессионных и

² Помимо авторов в исследовательских работах принимали участие Г. Т. Берсенева, Г. П. Бондаренко, И. П. Виленкина, Л. И. Зырянова, Г. Н. Кичигина, С. Д. Кулакова, Г. А. Нарышев, В. П. Полева, Т. Р. Сазонова, Д. С. Шпильфойгель. При проведении исследований использованы алгоритмы и программы, разработанные сотрудниками ИЭиОП СО АН СССР А. В. Беккером, Л. В. Говорской, М. Л. Лукацкой, Б. Г. Миркиным, Г. П. Ульяновой, Л. В. Цепляевой, Л. В. Черным, М. А. Ягольщицером.

балансовых зависимостей; дана постановка задачи многоцелевой оптимизации и обобщены методы ее решения; найдены подходы к построению и оптимизации производственных функций микропроцесса.

В межзаводских исследованиях объектом моделирования, как правило, были показатели эффективности функционирования целого промышленного предприятия. В статистическую совокупность объединялись предприятия одного или нескольких экономических районов или природно-географических зон, используемая информация носила смешанный, пространственно-временной характер, со значительным усреднением по временным интервалам (единицы наблюдения — месяц, квартал или год). Для подобных задач уже возникают проблемы, связанные с изучением структуры статистической совокупности, т. е. возможно использование моделей дискретно-непрерывного типа. Кроме того, методические трудности здесь вызывает также измеримость показателей и факторов, сопоставимость данных (из-за разного уровня агрегирования, региональных различий в ценах и т. п.). Такого рода исследования проводились по лесозаготовительной, угольной, цементной и электротехнической промышленности.

Информация о работе леспромпхозов Урала, Западной и Восточной Сибири изучалась за десятилетний период. На основе совокупности пространственных моделей, построенных за каждый год, установлены закономерности влияния природно-географических и технико-организационных факторов на экономические показатели лесозаготовки. Это послужило базой для определения экономической эффективности природных условий, соответствующего ей районирования лесосырьевых баз и для разработки системы групповых нормативов стимулирования леспромпхозов.

В угольной промышленности проводились исследования трех типов: построение, во-первых, пространственных моделей технико-экономических показателей угольных шахт для Донецкого, Кузнецкого, Карагандинского и Подмосковского бассейнов; во-вторых, динамических моделей работы отдельных угольных карьеров Кузбасса; в-третьих, моделировался процесс освоения производственных мощностей на угольных шахтах страны.

Цель исследования первого типа — выявить закономерности влияния горно-геологических и технико-организационных факторов на экономические показатели производ-

ства. В исследованиях второго типа обрабатывалась методика построения динамической модели отдельного объекта как элемента общей динамической модели совокупности объектов. Подобные модели предназначены для создания АСУ предприятия. К этому типу примыкают и задачи, решенные по информации о работе цементных заводов Урала, Сибири и Дальнего Востока.

Исследования третьего типа проводились для разработки методики прогнозирования параметров освоения проектной мощности вновь пускаемого объекта по комплексу априорно известных характеристик. Изучалась информация о работе 333 угольных шахт, вновь введенных в эксплуатацию. Полученные результаты представляют собой интерес для решения динамических отраслевых задач.

Выполнено два исследования по показателям строительства промышленных объектов. В первом из них устанавливалась зависимость удельных капитальных вложений от объема производства и сортамента продукции; во втором строились типичные кривые распределения капиталовложений по этапам строительства предприятий химической промышленности. Оба исследования содержат решение типичных задач, которые легко приложимы к условиям строительства предприятий других отраслей промышленности. При решении второй задачи найден методический подход к построению статистических моделей при наличии априорных ограничений на область допустимых значений моделируемого показателя и факторов.

Несколько иной постановкой отличается исследование по электротехнической промышленности. Здесь на первый план выступает многоименность продукции, и задачей экономико-статистического моделирования было установить влияние характеристик вида продукции (электрогенераторы и электронные лампы) на удельные расходы материалов и себестоимость.

В задачах этой группы особый интерес представляют исследования, объектом которых выступает совокупность всех предприятий определенной отрасли. В этом случае также моделируются экономические показатели эффективности функционирования предприятия в целом или его крупных подразделений, информация носит смешанный, пространственно-временной характер. Однако характер вариации моделируемых показателей и производственных факторов особый. Здесь можно выделить два момента.

Во-первых, вариационный размах значений показателей и факторов практически является максимально возможным для условий нашей страны. Во-вторых, в полной мере начинают сказываться качественные различия между объектами статистической совокупности. С этим связаны как определенные преимущества, так и особые трудности. Преимущества вытекают из глобального характера исследования, когда появляется возможность анализировать пространственную структуру отрасли и ее развитие во времени. Трудности же обуславливаются наличием неоднородности статистической совокупности, требующей разработки особых подходов к экономико-статистическому моделированию.

Задачи подобного типа решались по мартовскому производству, угольным шахтам и цементным предприятиям. В ходе исследований отработаны вопросы учета влияния цен, соизмерения признаков различного типа и уровня агрегирования, выделения однородных подсовокупностей, анализа устойчивости типологической структуры отрасли во времени и т. д.

Научные и прикладные результаты проведенных исследований получили отражение в девяти сборниках научных трудов ИЭиОПП СО АН СССР, посвященных этой проблематике; в целом ряде статей, опубликованных в журналах; в выступлениях и докладах на научно-технических конференциях и симпозиумах, а также в следующих монографиях: «Экономико-статистические исследования промышленного производства» (М., «Статистика», 1969); «Теория распознавания образов в экономических исследованиях» (М., «Статистика», 1973); «Распознавание образов при построении экономико-статистических моделей» (Новосибирск, «Наука», 1975); «Статистические модели в оптимальном отраслевом планировании» (М., «Статистика», 1975); «Статистическое моделирование экономических показателей» (Новосибирск, «Наука», 1976). Рассмотренные в них вопросы экономико-статистического моделирования промышленного производства и применения ЭСМ в экономической работе обобщены и систематизированы в настоящей публикации.

При написании данной книги учитывалось ее назначение — служить методическим пособием для специалистов, занимающихся применением статистических методов в экономических исследованиях промышленного производ-

ства. Этим определяется структура работы, характер и стиль изложения.

В первом разделе книги рассмотрены общеметодические вопросы организации экономико-статистического исследования и построения ЭСМ. Большое внимание уделено двум слабо разработанным проблемам экономико-статистического моделирования: неоднородности совокупности и учету ее типологической структуры, а также отражению в моделях динамики развития объектов изучаемой совокупности. При решении этих проблем широко используется новый (для экономических исследований) аппарат теории распознавания образов.

Если применение методов корреляционно-регрессионного анализа в экономических исследованиях стало уже в некотором смысле традиционным, то приложение методов теории распознавания образов к экономическим задачам представляет собой принципиально новый подход. Он существенно обогащает арсенал средств исследования промышленного производства и позволяет решить ряд актуальных проблем экономико-статистического моделирования. Одна из таких проблем — построение адекватной модели промышленного объекта. Дело в том, что промышленные предприятия — достаточно сложные объекты, объем априорной информации о закономерностях их функционирования недостаточен, исходные совокупности отличаются неоднородностью и сравнительной малочисленностью, большую роль играют качественные признаки.

Перечисленные особенности экономических совокупностей затрудняют непосредственное применение к их анализу «классического» аппарата математической статистики. Предлагаемые в работе математические подходы основаны на синтезе идей и методов корреляционно-регрессионного анализа и теории распознавания образов и позволяют преодолеть указанные затруднения и расширить область применения методов экономико-статистического моделирования. Аппарат теории распознавания образов используется как на стадии предмодельного анализа (группировка признаков и объектов, анализ однородности совокупности, обработка материалов экспертного опроса и пр.), так и непосредственно при построении моделей разного типа (классификационные, статические и динамические дискретно-непрерывные модели и т. д.).

Второй раздел книги посвящен вопросам применения ЭСМ в анализе, нормировании и планировании промышленного производства. Он содержит комплекс специальных прикладных методик решения типичных задач внутри- и межзаводского уровня. В гл. 5 большое внимание уделено двум задачам технического нормирования и внутризаводского планирования — оценке и прогнозированию переводных коэффициентов и выбору экономически выгодного режима эксплуатации крупного технологического агрегата.

Предлагаемые варианты методики установления переводных коэффициентов (коэффициенты производительности различных видов оборудования, трудоемкости производства отдельных видов продукции, коэффициенты для корректировки фондов заработной платы и т. д.) практически охватывают большинство типичных ситуаций планово-нормативной практики. Они длительное время используются на ряде металлургических предприятий, вошли в общепромышленные положения по техническому нормированию в черной металлургии и применяются также на предприятиях угледобывающей промышленности.

В гл. 6 и 7 рассматриваются вопросы моделирования основных этапов развития как отдельного промышленного предприятия, так и использования ЭСМ в межзаводском анализе и планировании, показано значение статистических моделей предприятий в отраслевом планировании.

В приложениях излагаются алгоритмы программ, которые использовались на различных этапах экономико-статистических исследований: предмодельном, построения и анализа статических и динамических моделей.

Используемый в работе аппарат математической статистики и теории распознавания образов не излагается, даны ссылки на соответствующую литературу. Рекомендации рассчитаны на читателей, знакомых с основами математической статистики и теории распознавания образов.

В написании работы принимали участие: Б. Б. Розин (введение, гл. 1—5, гл. 6, § 1, 2, гл. 7, § 2, 3); М. Л. Лукацкая (введение, гл. 3, 4, приложения 1, 3, 4); Н. М. Журавель (введение, гл. 2, 3, 5, 7, § 1, приложение 1); Л. А. Сергеева (гл. 1, 6, § 1, 2, гл. 7, § 3); М. А. Ягольничер (гл. 3, приложения 1, 2); Н. В. Овчинникова (гл. 7, § 2, 3); Н. П. Ионина (гл. 2, § 3, 4, гл. 7, § 1); В. В. Кулешов (гл. 7, § 4); Л. П. Талышева (гл. 6, § 3, приложение 3). Список литературы подготовлен Н. В. Овчинниковой.

Авторы приносят искреннюю благодарность докт. экон. наук В. И. Чепланову, Л. Д. Юдиной (Центральный научно-исследовательский институт черной металлургии, г. Москва), кандидатам экон. наук В. Н. Лушникову, А. С. Решетникову, В. П. Стяжкину (Научно-исследовательский и проектно-конструкторский институт лесной и деревообрабатывающей промышленности, г. Тюмень), канд. экон. наук М. Ц. Мкртчяну (Центральный научно-исследовательский институт комплексного электропривода, г. Ереван), М. Д. Зейгеру (Сибирский научно-исследовательский институт цементной промышленности, г. Красноярск), докт. техн. наук Г. Д. Мигачеву (Гипроуглеавтоматизация, г. Москва) за помощь и консультации при решении отраслевых задач; докт. техн. наук Н. Г. Загоруйко, канд. физ.-мат. наук В. Н. Елкиной и Г. С. Лбову (Институт математики СО АН СССР, г. Новосибирск), докт. техн. наук Н. М. Браверману, канд. техн. наук И. Б. Мучнику (Институт проблем управления АН СССР, г. Москва) за предоставление алгоритмов и программ и консультации при их использовании.

Раздел I

МЕТОДОЛОГИЧЕСКИЕ И МЕТОДИЧЕСКИЕ ВОПРОСЫ ЭКОНОМИКО-СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Глава I

ОСНОВНЫЕ ПОЛОЖЕНИЯ

§ 1. Объект и предмет исследования

1.1. В каждом исследовании необходимо четко определить, какую проблему предстоит решать и что составляет объект и предмет исследования. Ответ на первый вопрос дан во введении. В настоящем параграфе рассматриваются объект и предмет исследования, на основе чего в последующем определяется содержание задачи статистического моделирования.

1.2. Объектом исследования являются обособленные звенья того уровня организации системы промышленного производства, на котором непосредственно осуществляется материальный процесс преобразования ресурсов в продукцию, — технологические агрегаты, цехи, предприятия. Вышестоящие звенья (уровни) системы промышленного производства (тресты, комбинаты, главки, министерства) непосредственно не связаны с материальным процессом производства, а осуществляют переработку информации о работе промышленных объектов и управляют ими на основе полученных результатов.

1.3. Исследованию и моделированию обычно подлежат не сами по себе промышленные объекты, а процессы их функционирования. Следовательно, непосредственно предметом моделирования выступают не сами объекты с их физическими, химическими и другими конструктивными особенностями, а их свойства, отраженные в величинах основных показателей и факторов функционирования и во взаимозависимостях, существующих между ними.

1.4. Для более четкого представления о предмете статистического моделирования необходимо рассмотреть

структуру и схему производственного процесса на промышленном объекте. Они достаточно подробно описаны в литературе (см., например: Розин, 1976; Хейман, 1964; Реусс, 1963; Экономико-статистические исследования промышленного производства, 1969; и др.), и здесь мы дадим им лишь самую краткую характеристику.

Все структурные элементы промышленного производства (рабочая сила, материально-вещественные элементы: орудия, предмет и продукт труда, технология и организация производства) взаимосвязаны и образуют не произвольные, а определенные типичные комбинации. Связи между ними не носят жесткого детерминистского характера, неоднаправленны и неоднозначны.

В технико-экономических исследованиях работа промышленного объекта рассматривается как процесс преобразования предмета труда в продукт труда, чему сопутствует изменение и состояния самого предприятия.

Так как в ходе исследования мы всегда имеем дело только с информацией о материальных потоках в виде технико-экономических показателей, то производство можно рассматривать как преобразователь входной информации в выходную. Отсюда все характеризующие деятельность промышленного объекта показатели можно подразделить на входные, условия производственного потребления ресурсов и выходные (результаты производства).

Между выходными показателями, с одной стороны, и всеми остальными, с другой — существует некая закономерная связь. Имея модель такой связи, можно вести экономические расчеты и управлять выходными показателями.

Для целей статистического моделирования удобно принять условное деление всех характеристик процесса производства на производственные факторы и результативные экономические показатели¹.

Выделим два вида оценок функционирования промышленного объекта — интенсивность и эффективность. В экономической интерпретации интенсивность производства для предприятия — это объем выпуска продукции за год (или другой период), для агрегата — производительность в час,

¹ В дальнейшем изложении будут употребляться наряду с терминами «производственные факторы» и «результативные экономические показатели» соответственно «факторы производства» и «экономические показатели», или просто «показатели».

смену и т. д. Эффективность производственного процесса есть отношение выходного потока к входному, соотношение результатов и затрат производства. (Возможна и обратная форма такого показателя, когда затраты рассчитываются на единицу выпускаемой продукции.)

Для уровня предприятия следует различать прямую (локальную) и народнохозяйственную (глобальную) эффективность затрат. Различие этих видов эффективности соответствует различиям в действительных, прямых и дифференциальных затратах (Новожиллов, 1972). В настоящем исследовании моделированию подлежат показатели прямой (локальной) эффективности производственного процесса без учета затрат обратной связи.

1.5. Итак, в экономико-статистическом исследовании изучается процесс формирования показателей интенсивности работы промышленного объекта и эффективность протекания производственного процесса. При этом промышленный объект рассматривается как автономный, внешние условия функционирования которого заданы системой более высокого уровня.

§ 2. Задача и основные этапы экономико-статистического моделирования

1.6. Экономико-статистическая модель — одна из форм экономико-математических моделей. Она строится на основе обобщения массового эмпирического материала, предназначена для раскрытия закономерностей, проявляющихся в совокупности явлений, и включает элементы неопределенности. Наиболее специфическая черта ЭСМ состоит в том, что она выступает как средство описания стохастических связей и закономерностей, возникающих под воздействием множества причин и следствий в массовых, повторяющихся явлениях.

Возможность экономико-статистического моделирования процесса функционирования экономического объекта основана на признании того факта, что связи между экономическими показателями и производственными факторами по своей природе стохастические. Вполне понятно, что это замечание не относится к составным элементам расчетной формулы показателя. Их связь с величиной

показателя функциональная. Такие факторы можно назвать «прямыми»², в отличие от опосредованных («стохастических»), определяющих вариацию уже рассчитанных величин показателей.

Статистическая модель функционирования объекта состоит из уравнений и неравенств, параметры которых оценены методами математической статистики. Наиболее часто такую ЭСМ составляют регрессионные уравнения (одно или несколько) и комплекс вспомогательных статистических характеристик (средние, дисперсии, коэффициенты корреляции, оценки точности аппроксимации и т. д.). Используются и другие формы экономико-статистических моделей (например, классификационные, дискретно-непрерывные).

1.7. В образованной любым способом статистической совокупности объектов имеет место вариация экономических показателей. С точки зрения моделирования ее можно подразделить на две части: «объяснимую», вызываемую изменением известных и поддающихся учету в модели факторов производства и «остаточную» — результат воздействия не поддающихся учету, измерению и частично неизвестных факторов. Кроме влияния неучтенных факторов производства остаточная вариация есть следствие неадекватности формы модели реальному процессу и неточностей и погрешностей в измерении величин экономических показателей.

Причины объяснимой вариации экономических показателей следующие:

- 1) различие масштабов производства при фиксированной комбинации его структурных элементов;
- 2) вариантность комбинаций структурных элементов при постоянном уровне производства;
- 3) качественные различия самих структурных элементов производства, возникающие в основном под воздействием технического прогресса;
- 4) неодинаковый уровень использования ресурсов;
- 5) разные «возраст» и момент пуска предприятий, т. е. последние исследуются на различных точках траектории своего развития;

² «Прямые» факторы тоже нельзя точно оценить. Однако существенное различие между ними и стохастическими факторами, конечно, есть.

б) влияние на функционирование объектов неодинаковых условий внешней среды.

Основное содержание задачи статистического моделирования объектов и состоит в оценке меры участия производственных факторов в объяснимой вариации исследуемого показателя. В геометрической интерпретации это означает, что устанавливается зависимость между положениями точек в пространстве экономических показателей и в пространстве факторов производства. Для удобства используется еще понятие перемещения объектов в признаковых пространствах, и тогда содержание задачи моделирования сводится к описанию связи перемещения точки (объекта) в пространстве экономических показателей с ее перемещением в пространстве факторов производства.

Формально содержание задачи экономико-статистического моделирования можно записать следующим образом. Известны вектор результативных экономических показателей $Y = (y_1, y_2, \dots, y_m)$ и вектор производственных факторов $X = (x_1, x_2, \dots, x_n)$. Необходимо построить экономико-статистическую модель объекта, а именно оператор $F(x)$, который устанавливает соответствие между входными и выходными переменными и дает возможность с определенной точностью восстановить экономические показатели объекта по значению его производственных факторов. Понятно, что зависимость \vec{Y} от \vec{X} можно описать с помощью бесконечного множества функций. Для выбора вида функции надо знать сам моделируемый процесс и учитывать возможности используемого математического аппарата.

4.8. В организации экономико-статистического исследования выделяются следующие этапы: экономическая постановка задачи; выбор и построение рационального типа модели, анализ и экономическая интерпретация результатов, разработка предложений по использованию модели.

Постановка задачи заключается в определении общих путей и методов реализации поставленной цели, для чего должен быть выбран способ формирования совокупности, набор переменных, включаемых в анализ, методы статистического анализа и моделирования.

Отдельные этапы экономической постановки задачи детально описываются в гл. 2.

Выбор типа модели для описания анализируемого процесса представляет собой один из наиболее важных моментов экономико-статистического исследования. Он

должен осуществляться с учетом как общих требований к статистической модели, так и специфических условий проводимого конкретного исследования. К общим требованиям следует отнести:

а) наиболее полное отображение реальной структуры исследуемой совокупности объектов в соответствии с принципом структурного подобия;

б) максимально возможное использование информации, содержащейся в исходных данных;

в) как можно большую простоту модели, обеспечивающую решение задачи с достаточной точностью.

Эти общие требования к типу модели зачастую оказываются противоречивыми, и тогда между ними следует найти наиболее рациональный компромисс.

Специфичны для каждого конкретного исследования назначение будущей модели, качественные признаки в описании объекта, объем исследуемой совокупности и т. п.

Наконец, укажем еще одно обстоятельство, влияющее на выбор типа модели, — это возможности имеющихся в распоряжении исследователя алгоритмов, программ и вычислительной техники.

4.9. Построение математической модели — формальный этап, если задача поставлена, подготовлена необходимая информация, определены тип и форма модели, выбран алгоритм. Наиболее сложным на этом этапе представляется выбор подходящего алгоритма построения модели. В основе такого выбора лежат как содержательные, так и организационные соображения. К первым относятся подбор алгоритма, обеспечивающего достаточную точность модели, к организационным — возможность реализации того или иного алгоритма (наличие программы, соответствующей ЭВМ, трудоемкость подготовки и обработки информации и т. д.). Выбор алгоритма существенно зависит также от типа решаемой статистической задачи и назначения модели. Процедура выбора и построения рационального типа статистической ЭСМ подробно описана в гл. 3, динамической — в гл. 4. *в соответствующих главах*

1.40. Анализ и интерпретация результатов — полностью неформальный и один из наиболее сложных и ответственных этапов. Для интерпретации нельзя дать никаких алгоритмов или рецептов. Она полностью зависит от профессиональных и математических знаний исследователя. Единственное жесткое ограничение заклю-

чается в том, что интерпретация должна согласовываться с первоначально выдвинутыми гипотезами. Если такого согласования нет, гипотезы меняются и задача решается с самого начала.

Статистическая модель, даже самая совершенная, не способна охватить, выразить или отразить явления и процессы во всем их диалектическом единстве и многообразии. Любая модель представляет реальное явление только в упрощенном виде, отражая наиболее важные (с точки зрения целей исследования) его характеристики и связи.

Интерпретация и применение экономико-статистической модели всегда будут содержать элемент риска (если относительно неучтенных факторов можно сделать объективные вероятностные суждения) или неопределенности (если вероятности не могут быть определены). Обычно при использовании ЭСМ имеет место комбинация обоих указанных видов погрешностей с преобладанием в зависимости от конкретной ситуации одного из них.

1.11. Основное условие эффективного применения статистической модели — ее адекватность реальному явлению, процессу. Неадекватность ЭСМ может быть обусловлена тремя причинами: информационной (погрешности информации, ее недостаточность по набору показателей и факторов или по объему совокупности), структурной (наличие невыявленных неоднородностей) и модельной (использование неправильной формы зависимости между показателями и факторами).

Трудно заранее, особенно при прогнозировании, построить точные формальные критерии адекватности ЭСМ. Для оценки адекватности часто пользуются известным паллиативом — ретроспективной проверкой точности модели по известным статистическим критериям (Айвазян, 1968; Браунли, 1949; Ван дер Варден, 1960; и др.). Однако в конечном счете главный критерий адекватности отображения реально происходящего процесса созданной моделью — это практика приложений экономико-статистической модели. Только она позволяет судить о степени адекватности модели или о необходимости ее уточнения, корректировки.

Требования к адекватности модели зависят от ее типа. Можно различать модели «пассивного» типа, предназначенные для анализа и пассивного прогноза, и модели «активного» типа — используемые для управления. В первом

случае допускается соответствие между явлением и моделью только в одну сторону (явление → модель). Модель управления должна обеспечивать двухстороннее соответствие (явление → модель → явление) и адекватно отражать поведение экономической системы, а не только ее структуру. При изменении входов в модели должны быть такие же изменения выходов, как и на моделируемом объекте.

При прогнозе и принятии решений модель прилагается к совокупности, отличной от исходной, с другими пространственно-временными границами. Здесь понятие адекватности модели отображаемому явлению связано с понятием устойчивости отображаемых ею связей, приложимости модели в другие периоды времени и к другим объектам данной отрасли.

§ 3. Классификация экономико-статистических моделей

1.12. Можно выделить два главных момента, определяющих вид будущей модели функционирования промышленного объекта: назначение модели, экономический заказ к ней и способ реализации этого заказа. Оба момента взаимосвязаны, но не жестко. Существует определенная свобода в сочетании экономического заказа и способов его реализации.

Назначение модели для прикладного планово-экономического использования — понятие комплексное. Оно должно включать характеристику функциональной цели построения модели; уровень административно-хозяйственного органа, для которого модель предназначена, интервал времени, в котором предполагается ее использование. Назначение модели, понимаемое в вышеуказанном смысле, определяет и конкретные требования к ней, набор ее характеристик. В качестве характеристик, выступающих как классификационные признаки назначения модели, можно назвать следующие:

- набор (вектор) экономических показателей, подлежащих моделированию;
- масштаб исследования и уровень объекта моделирования;
- степень учета динамических закономерностей;
- набор переменных модели;
- степень общности модели.

1.13. По видам моделируемых экономических показателей следует различать модели показателей интенсивностей потока преобразования ресурсов в продукцию (объем производства) и показателей эффективности процесса преобразования различной степени общности (производительность труда, фондоотдача, удельный расход материалов, себестоимость, рентабельность и т. д.).

Показатели первого типа могут моделироваться для промышленных объектов любого уровня. Из группы эффективности производства для объектов более низкого уровня системы промышленного производства (например, агрегат) моделируются обычно только частные показатели (удельные расходы какого-либо ресурса).

По этому же классификационному признаку следует различать модели одного показателя и модели, описывающие процесс формирования вектора показателей.

1.14. Масштаб исследования определяет выделение двух типов моделей — внутри- и межзаводского анализа и планирования. Первые предназначены к использованию в пределах отдельного предприятия, вторые — для анализа и принятия решения по группе промышленных объектов. Это деление отчасти коррелирует и с уровнем объекта моделирования.

1.15. По характеру отражения динамики процесса различают статические и динамические модели. Первые охватывают единичный, фиксированный интервал времени, динамические модели отражают поведение системы в последовательные интервалы времени. Хотя модели, построенные по пространственному сечению, статичны, однако совокупность их за ряд лет позволяет делать определенные выводы о движении системы. При этом создается возможность динамизации пространственных моделей путем рассмотрения динамики их параметров, сдвига факторов во времени и т. д.

1.16. В зависимости от набора переменных модели могут быть полными и частными. Полные включают весь набор измеримых признаков, частные — наборы, характеризующие одну сторону процесса производства, например природно-климатические условия. Сравнением полных моделей с частными определяется степень влияния определенного набора признаков.

1.17. По степени общности различают автономные и системные модели экономических показателей. Первые предназначены для самостоятельного использования, модели второго вида входят органической составной частью в какую-либо систему моделей, что, естественно, налагает на них определенный комплекс требований и определяет характер их применения.

Этот же признак классификации служит основанием для деления моделей на одноуровневые, когда модель экономического показателя строится по схеме «производственные факторы → показатель экономической эффективности», и ступенчатые, многоуровневые, формируемые по схеме «производственные факторы → частные результирующие показатели → общие показатели экономической эффективности». Вторая схема в ряде случаев может оказаться предпочтительнее, так как она позволяет учесть большее число первичных факторов производства и проследить их влияние на общие синтетические показатели эффективности через посредство частных технико-экономических показателей.

Можно также различать общие модели для отрасли (подотрасли) и систему локальных моделей для отдельных объединений (регионов). Локальные более точны, чем общие. Дополнив общую модель системой локальных, можно проанализировать связи экономического показателя в пространственно-региональном разрезе.

1.18. Классификация моделей по способу построения многовариантна. Практически на каждом этапе построения модели имеется большая возможность выбора подходов и методов решения. Разные сочетания этих вариантов формируют большое количество видов моделей по способу их построения. Остановимся на классификации по трем признакам: способу описания анализируемых зависимостей; характеру их модельного описания и виду используемой исходной информации.

1.19. Различают три способа описания зависимостей: два так называемых «чистых» — при помощи классификационных (дискретных) или регрессионных (непрерывных) моделей и один комбинированный — при помощи дискретно-непрерывных моделей.

Регрессионная (непрерывная) модель для всей совокупности позволяет в явной форме оценить влияние отдельных производственных факторов, но она непригодна для

неоднородных совокупностей, так как в этом случае не обеспечивает соблюдения принципа структурного подобия.

Классификационная (дискретная) модель охватывает две классификации (в пространстве экономических показателей и факторов производства) и описывает связи между ними. Она используется для моделирования структурно-неоднородных совокупностей.

При дискретно-непрерывном моделировании интересующая нас зависимость записывается комбинацией моделей двух типов: дискретной, описывающей типологическую структуру совокупности, и системы непрерывных моделей объектов внутри классов.

Классификационные модели позволяют учитывать общее влияние комплекса признаков, при этом затруднений с учетом качественных признаков не возникает, нет ограничений, связанных с формой связи и однородностью совокупности, и т. д. Но в то же время здесь есть существенные потери информации за счет усреднения.

Анализ материалов по целому ряду исследований, показывает, что наиболее универсален комбинированный тип модели — дискретно-непрерывный. При рациональном сочетании составных частей дискретно-непрерывная модель более адекватна исследуемому процессу, чем модели, получаемые каждым из «чистых» способов моделирования в отдельности.

1.20. По характеру описания зависимостей можно выделить два типа ЭСМ — описательные и объяснительно-прогнозные. Описательные наилучшим образом отражают взаимосвязи переменных. Параметрам модели не приписывается определенного содержательного смысла, при их оценивании решается задача аппроксимации. Объяснительно-прогнозные модели устанавливают соответствие между совокупностью выявленных факторов и гипотезами, которые их включают или по крайней мере делают очень вероятными.

Здесь в основе отбора факторов-аргументов лежит изучение причинного механизма формирования прогнозируемого показателя, т. е. выявление движущих сил развития промышленного объекта. При оценивании параметров объяснительно-прогнозных моделей решается задача идентификации. Она заключается в том, чтобы с помощью подходящих статистических методов найти неизвестные пара-

метры в уравнениях, построенных на основе содержательных предположений.

1.21. Классификация моделей по виду исходной информации соответствует классификации по изучаемому виду вариаций. Следует различать три вида вариации экономических показателей: отдельного объекта во времени, совокупности объектов в пространстве и общую вариацию показателей совокупности объектов в пространстве и во времени.

Первый вид можно представить как изменение положения определенного объекта (изображающей его точки) в пространстве экономических показателей в последовательные периоды времени. При моделировании одного экономического показателя такое пространство есть плоскость, траектория движения точки представляет собой динамический ряд.

Второй вид вариации можно ассоциировать с разбросом в расположении точек, соответствующих различным объектам в фиксированный период времени. В случае моделирования одного экономического показателя будем иметь ряд распределения.

Вариация третьего вида — это комбинация двух первых и может трактоваться как дискретный случайный процесс. При такой трактовке динамику показателей отдельного предприятия можно считать одной реализацией такого процесса, а вариацию показателей совокупности объектов в определенный период времени — пространственным сечением процесса.

Общая вариация показателей формируется двумя способами: как общность временных рядов показателей объектов, входящих в совокупность, и как движение рядов распределения показателей объектов совокупности во времени. Выбор способа описания зависит от постановки задачи исследования и характера исходной информации.

1.22. Соответственно выделенным видам вариации различаются три типа статистических моделей экономических показателей.

Частная динамическая модель отдельного объекта связывает временное движение определенной точки в пространстве экономических показателей с изменением положения данного объекта в пространстве факторов производства. Она преимущественно используется для внутривзаводского анализа, нормирования и управления.

Частная статическая модель совокупности объектов объясняет различное положение предприятий в пространстве экономических показателей различием их положений в пространстве факторов производства в фиксированный период времени. Эта модель обычно строится для уровня предприятия (цеха) и используется для аналитических целей на более высоких уровнях управления (главк, министерство).

Общая динамическая модель вариации экономических показателей в совокупности объектов оценивает влияние на экономические показатели произвольных перемещений в пространстве факторов производства (под произвольными понимаются возможные перемещения между заданным конечным набором точек, описывающим состояние исследуемой совокупности объектов в разные периоды времени). Общая динамическая модель в отличие от статической может использоваться не только для анализа, но и для прогноза и принятия решений по охваченной исследованной группе объектов. Методика построения рассмотренных типов моделей изложена в гл. 3 и 4.

§ 4. Основные направления применения ЭСМ³

1.23. В экономических исследованиях ЭСМ находят разностороннее применение. Так, они выступают достаточным эффективным инструментом совершенствования методики экономического анализа, давая значительно большую количественную информацию, чем традиционные методы анализа. С помощью ЭСМ можно оценить зависимость результатов производства от таких входных факторов, влияние которых другими методами установить довольно трудно.

В технико-экономическом анализе экономико-статистические модели используются в основном для выявления внутривладельческих резервов и путей их использования, для объективной оценки хозяйственной деятельности.

Известны два направления аналитической работы на основе моделей: анализ результатов деятельности отдельного объекта во времени (сравнение с прошлым периодом, с планом, выявление факторов, определяющих динамику

показателя) и сравнительный анализ деятельности совокупности промышленных объектов. Первое направление (его можно назвать динамическим анализом) преследует такие цели: определить тенденции, дать оценку технического и экономического развития объекта, выявить важнейшие резервы улучшения показателей для данного объекта. Одна из основных целей сравнительного анализа — выявление экономически выгодных вариантов производства из серии реализованных в совокупности.

1.24. При использовании статистических моделей в целях прогнозирования неизбежен выход за временные и пространственные рамки той информации, по которой построена ЭСМ. Принципиальная возможность распространения выводов выборки на будущий период основана на определенной устойчивости технико-экономических закономерностей. Однако в каждом конкретном случае возможность экстраполяции должна быть логически проверена.

Модель прогноза относится к имплицативному типу и строится по схеме: «если, то» (Вишнев, 1968). Она пригодна в пределах того периода времени, в котором моделируемый способ производства не претерпевает коренных изменений; практика показывает, что этот срок неодинаков для разных отраслей и объектов. За его пределами речь может идти только о более агрегированном прогнозе — тенденций, пропорций и других укрупненных характеристик.

Различают два подхода к прогнозированию (Янч, 1970). Во-первых, от сегодняшнего дня, от имеющегося базиса знаний и ресурсов, постепенно проникая в будущее. Во-вторых, сначала определив будущие цели и ориентиры, осознав результаты, которых желательно достичь в будущем, уже от них постепенно двигаются к настоящему, увязывая желания с возможностями. Первый способ называется изыскательским, исследовательским прогнозированием, второй — нормативным. Статистическое прогнозирование по своей сути исследовательское. Оно независимо, непредвзято, но имеет и недостатки: пассивность, некоторую инертность по отношению к сложившимся системам техники и экономики.

Статистическое прогнозирование экономических показателей представляет собой многошаговую процедуру и в общем случае включает следующие этапы:

конструирование модели прогноза;

³ Этот вопрос на конкретных примерах подробно рассматривается во втором разделе книги.

оценку значений независимых переменных модели на момент прогнозирования и собственно прогноз;

анализ точности прогнозирования и систематическую его корректировку.

Собственно прогноз можно записать в точечной или интервальной форме.

В первом случае прогнозируется математическое ожидание случайной величины ($M[Y]$), во втором — интервал, к которому с некоторой вероятностью должна принадлежать прогнозируемая величина в определенный момент времени, т. е. указывается интервал ΔP , для которого $P(Y \in \Delta P) = P$.

Удобна и комбинация этих форм прогноза: ожидаемое значение прогнозируемого показателя записывается в виде $a \pm s$, где a и s — соответственно оценки математического ожидания и среднего квадратического отклонения. Заметим, что прогнозировать можно как одномерную, так и многомерную случайную величину.

По характеру принимаемых решений следует различать пассивный и активный прогноз. На основе первого принимаются решения лишь с учетом прогнозируемых показателей (планово-нормативные задачи), на основе второго — решения, направленные на изменение, регулирование этих показателей (задачи по принятию оптимальных плановых решений и управлению производством).

1.25. В планово-нормативных расчетах ЭСМ могут использоваться при учете объективных условий производства, экономической оценке технического уровня производства и природных условий.

Сущность первой задачи состоит в оценке степени влияния на вариацию экономических показателей различных неуправляемых объективных условий производства. Выделение и соизмерение таких условий позволяет установить некоторый диапазон значений, которые может принять экономический показатель при средних значениях усилий производственного коллектива. В этом случае соизмерить объективные условия производства — значит сравнить числа, характеризующие этот диапазон, со средним по всей совокупности значением. Полученные таким образом коэффициенты-соизмерители объективных условий производства могут быть широко использованы в экономических приложениях, связанных с обоснованием и распределением плановых заданий, обоснованным прогнозом

объективных возможностей предприятий, установленном нормативов экономического стимулирования, групповых расчетных цен и т. д.

Вторая задача, а именно оценка экономических последствий технического прогресса, особенно важна на современном этапе развития промышленного производства. Действие технического прогресса проявляется в совершенствовании структурных элементов промышленного производства и в изменении экономических показателей производства. Экономико-статистические модели объектов производства позволяют относительно соизмерить технический уровень производства различных промышленных объектов и экономические последствия изменения этого уровня.

Экономическая оценка эффективности природных условий имеет особое значение для добывающих отраслей промышленности, где роль этой группы факторов в вариации экономических показателей весьма значительна. Полученные на основе ЭСМ относительные оценки эффективности отдельных природных факторов и их колебаний затем могут быть использованы для экономического районирования территории, но не общего, а специфического (с позиций интересов определенной отраслевой системы).

1.26. В оптимальном отраслевом планировании и ЭСМ находят применение для решения следующих задач.

Внутриотраслевой экономический анализ функционирования объектов отрасли проводится на стадии постановки отраслевой задачи и заключается в исследовании состояния отрасли, направлений ее развития, динамики основных показателей развития отдельных объектов и всей отрасли в целом. В результате определяются факторы, влияющие на экономические показатели функционирования объектов отрасли, на размещение предприятий и т. д.; перспективы развития отдельных объектов и всей отрасли.

В качестве информационных экономико-статистические модели применяются для прогнозирования значений параметров задач отраслевого планирования. В частности, могут быть рассчитаны объемы производства продукции, текущие и капитальные затраты и т. д. Использование ЭСМ значительно расширяет набор вариантов, сокращает объем вычислительных работ, повышает обоснованность показателей по вариантам.

Построение двухуровневой модели «предприятие — отрасль», в которой ЭСМ объекта выступает органической составной частью оптимизируемой системы, предполагает создание новой разновидности модели — оптимальных статистических моделей (подробно эти модели будут рассмотрены в гл. 5). Это позволяет существенно усовершенствовать процесс оптимизации на обоих уровнях отраслевой системы.

1.27. Наконец, ЭСМ применяются в управлении производственными процессами. Для принятия практических решений и управления реальными процессами пригодны только экономико-статистические модели «активного» типа, которые адекватно отображают поведение объекта при различных воздействиях и его реакцию на управляющие факторы. Кроме того, применяемые в управлении ЭСМ обязательно должны иметь элемент нормативности и включать критерии, характеризующие цель управления. Таким образом, экономико-статистические модели управления носят смешанный функционально-нормативный характер.

§ 5. Методы экономико-статистического исследования

1.28. До последнего времени для целей экономико-статистического моделирования в основном использовались два вида статистических методов — группировок (главным образом комбинационная группировка) и корреляционно-регрессионного анализа. В настоящем исследовании математический аппарат моделирования расширен за счет включения в него нового формального аппарата — методов теории распознавания образов. Такое комплексное применение статистических методов позволяет существенно расширить сферу их приложения.

1.29. В связи с широкой известностью методов группировки и корреляционно-регрессионного анализа кратко остановимся только на характеристике задач, методов и основных понятий теории распознавания образов⁴.

⁴ Подробное изложение проблем теории распознавания образов содержится в работах С. А. Айвазяна, З. И. Бежаевой, О. В. Староверова (1974); М. А. Айзермана, Э. М. Бравермана, Л. И. Рознозара (1970); Э. М. Бравермана (1966, 1970, 1971); Н. Г. Загоруйко (1966, 1971, 1972); Дж. Надь (1968); Г. С. Себестяна (1965); И. Т. Турбовича, В. Г. Гитиса, В. К. Маслова (1971) и других.

Аппарат распознавания образов может выступать и как самостоятельный инструмент построения модели экономического процесса. Его применение позволяет решить целый ряд острых проблем, таких как формирование отдельных совокупностей, выбор существенных признаков, учет влияния качественных признаков и скачкообразного характера развития объектов, оценка устойчивости зависимостей во времени и т. д. Кроме того, при использовании методов распознавания образов снимаются ограничения на форму связи. В то же время необходимо отметить, что теория распознавания образов как сравнительно молодая область знаний находится в стадии становления и имеет много нерешенных проблем. Одна из них — это неоднозначность решения. Разные алгоритмы и программы могут давать на одном и том же материале неодинаковые результаты. В распоряжении исследователя пока нет эффективных формальных критериев для оценки и сравнения алгоритмов и программ распознавания образов. Поэтому необходимо осторожное, критическое отношение к результатам формального анализа.

Теория распознавания образов имеет дело с множествами объектов произвольной природы⁵. Каждый из них описывается определенным набором характеристик (признаков), с помощью которых его можно отличать от других объектов соответствующего множества.

Распознавание образов — это принятие решения о принадлежности нового, ранее в процессе изучения не рассматривавшегося объекта к данному классу (образу) путем сравнения свойств этого объекта с уже известными и изученными. Под классом (образом) понимается множество объектов или явлений, объединенных некоторым общим свойством (близких по набору тех или иных признаков).

В теории распознавания приходится иметь дело с некоторым набором классов, с так называемым «алфавитом классов». Каждый из классов включает в себя некоторое число отдельных объектов.

Для формирования образов необходимо выработать механизм, при помощи которого сходные между собой объекты могли бы быть отнесены к одному образу. Коли-

⁵ Например, объектами могут быть совокупности изучаемых предприятий, буквы какого-нибудь алфавита, звуки речи и т. д.

чественная характеристика образа — это мера сходства между его явлениями, т. е. та характеристика, на основании которой явления группируются и составляют образ.

На практике используются три основных способа количественного измерения меры сходства: с помощью коэффициентов ассоциации (связи), корреляции и показателей расстояния (Ван дер Варден, 1960; Житков, 1970; и др.).

1.30. Несмотря на относительно большое число работ по классификации задач распознавания образов, единой общепринятой точки зрения здесь пока нет. В изложении этой проблемы мы придерживаемся взглядов Н. Г. Загоруйко (1966—1972), наиболее приемлемых для рассматриваемого в настоящей работе круга прикладных вопросов. Можно выделить следующие три типа основных задач распознавания образов: обучения распознаванию, сокращения (минимизации) описания, таксономии, или, как ее иногда называют, «самообучения».

Математическая постановка задачи обучения распознаванию образов следующая.

Дана обучающая выборка в виде последовательности наблюдаемых объектов. Для каждого объекта указан класс, к которому тот принадлежит; заданы элементы алфавита S , указаны признаки x , по которым эти элементы следует отличать друг от друга, и величина допустимых потерь R (например, процент ошибок на экзамене). Требуется найти такое решающее правило (критерий) δ , которое обеспечивало бы распознавание элементов S в пространстве X с потерями, не превышающими R .

Под решающим правилом здесь понимается некоторая процедура, позволяющая выбирать решения из конечного их множества. Экзаменом называется операция по распознаванию на основе материала обучения объектов, ранее не участвовавших в обучении, однако принадлежность которых к определенным классам заранее известна. На основании результатов экзамена может быть вынесено решение о качестве обучения. Достигнув приемлемого качества обучения, можно вести распознавание любых объектов, наблюдаемых в данной ситуации.

Задача обучения распознаванию образов может быть поставлена в детерминистской и вероятностной формах. Главное их отличие заключается в том, что в первом случае имеется полная определенность относительно принад-

лежности объекта к данному классу, тогда как во втором такой определенности нет и приходится говорить о вероятности принадлежности объекта к данному классу.

1.31. Задача минимизации описаний может быть сформулирована следующим образом: найти такое преобразование исходного пространства признаков в некоторое другое пространство, чтобы функция потерь существенно не увеличивалась, а размерность нового пространства была гораздо меньше исходного. В условиях предыдущей задачи ее можно формализовать так: найти подсистему признаков x' , которая обеспечила бы распознавание элементов алфавита S с потерями, не превышающими R . Поскольку информация, заключенная в данном наборе признаков, должна обеспечивать достаточно высокое качество распознавания, его называют информативной подсистемой признаков. Несмотря на кажущуюся простоту постановки, задача нахождения информативной подсистемы признаков, к сожалению, не имеет однозначного решения.

1.32. Суть задачи таксономии⁶ заключается в следующем: имеется множество объектов, которое необходимо разделить с помощью какого-нибудь решающего правила на заранее заданное или незаданное число классов. В терминах упоминавшихся выше обозначений задача таксономии формулируется так: множество реализаций, заданных в пространстве X , с помощью решающих функций δ (т. е. по определенному критерию «близости», «сходства») нужно разделить на такое количество и таких элементов алфавита S , чтобы потери не превышали заданной величины R . Содержательно это означает получение «наилучшего» разбиения исследуемой совокупности на образы (классы) с точки зрения выбранного критерия. «Наилучшим» считается разбиение, когда в один и тот же класс входят элементы, относительно «похожие» друг на друга.

Как правило, на практике приходится решать задачи, представляющие комбинации ранее указанных типов. Например, одновременно с обучением распознаванию образов необходимо найти подмножество информативных призна-

⁶ В литературе наряду с указанным термином широко используются также «автоматическая классификация», «обучение без учителя», «самообучение», «кластер-анализ» и др.

ков. Для этого задачу решают в несколько этапов, находя последовательно решения каждой из основных при фиксированном решении других задач.

1.33. Разработано большое число алгоритмов решения основных задач распознавания образов. Подробные и квалифицированные обзоры содержатся в работах А. А. Дорофеюка (1971), Г. Н. Житкова (1970), И. Г. Турбовича, В. Г. Гитиса, В. К. Маслова (1971) и других авторов.

Выбор того или иного алгоритма или их комбинации для решения конкретной задачи диктуется соображениями как содержательного, так и организационного характера (объем исходной информации, профессиональные гипотезы о характере закономерностей, цель исследования, наличие программ и возможность их реализации на ЭВМ данного типа, опыт и предпочтение исследователя и т. п.).

В наших исследованиях наиболее широко применялись алгоритмы и программы, разработанные в Институте математики СО АН СССР (В. Н. Елкиной, Н. Г. Загоруйко, Г. С. Лбовым), Институте проблем управления АН СССР (М. А. Айзерманом, Э. М. Браверманом, Л. И. Розоноэром, А. А. Дорофеюком, И. Б. Мучником) и в Институте экономики и организации промышленного производства СО АН СССР (А. В. Беккером, М. А. Ягольницером). Ориентация именно на эти алгоритмы обусловлена наличием творческих связей с указанными коллективами и тем, что пригодность таких алгоритмов подтверждена значительным числом экспериментальных расчетов.

Глава 2

ПОСТАНОВКА ЗАДАЧИ ЭКОНОМИКО-СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ИСХОДНОЙ ИНФОРМАЦИИ

2.1. К вопросам экономической постановки задач относятся:

формулировка целей и определение объекта исследования;

априорный теоретический анализ исследуемого процесса функционирования объекта, формирование гипотез о

наборе переменных, характере и структуре связей и т. д.;
проведение экспертного опроса и анализ его данных;
выбор измерителей переменных;
определение источников информации и способов ее сбора;

оценка достоверности и надежности исходных данных.

Решение перечисленных вопросов основывается на логических положениях прикладных наук, материалах предыдущих исследований, обобщении накопленного опыта — исследований и практики. Полученные результаты нельзя считать окончательными, они скорее носят характер гипотез, которые должны уточняться в ходе построения модели. Уточнение необходимо по ряду причин. Во-первых, внутренняя логика производственного процесса как основа для профессиональной постановки задачи полностью не известна. Существенность и характер связей отдельных факторов часто выясняются лишь в результате статистического анализа. Во-вторых, должны быть учтены возможности и затраты по подготовке информации. Далее следует соизмерить эффект от уточнения ЭСМ с величиной затрат на ее построение. Наконец, имеются важные ограничения, связанные с математико-статистическим аспектом исследования.

§ 1. Формулировка целей исследования.

Априорный анализ процесса функционирования объекта

2.2. Сформулировать цель исследования — значит конкретизировать общую задачу применительно к исследуемой ситуации: определить, какие экономические показатели, по каким объектам и для какого применения должны моделироваться. По существу, это есть формирование экономического заказа (требований) к будущей модели.

Например, исследование по мартеновскому производству проводилось для построения системы моделей формирования обобщающего экономического показателя работы мартеновского цеха — себестоимости стали и ее важнейших составляющих: затрат на заданное сырье и расходов по переделу. Модели предназначались для использования в экономическом анализе и планировании работы предприятий черной металлургии.

Выбор объекта моделирования сводится к отысканию единицы наблюдения и заданию пространственных и вре-

менных границ совокупности. Иначе говоря, определяется уровень объекта наблюдения (крупный технологический агрегат, цех, предприятие, комбинат) с сопоставимой структурой и временная длительность единичного акта производства (час, смена, сутки, месяц, квартал, год).

Например, возможны три уровня исследования мартеновского производства — мартеновские печи, цехи и предприятия в целом. Последний уровень исключается из-за большой разнотипности мартеновских цехов на многих металлургических предприятиях. Мартеновские печи также не могут выступать основным объектом экономико-статистического моделирования, так как не представляют собой самостоятельной хозяйственной единицы. В то же время они могут быть вспомогательным объектом в исследованиях экономического характера.

Объектом наших исследований были мартеновские цехи как сравнительно однородные структурные подразделения, находящиеся на самостоятельном балансе. Именно на цеховом уровне осуществляется учет всех затрат на производство и экономический анализ по сопоставлению затрат и результатов — оценка эффективности производства.

В экономических исследованиях время функционирования объектов выступает в дискретной форме (Розин, 1976). На выбор длительности временного отрезка (кванта) влияет уровень объекта, способ формирования совокупности, периодичность измерения и учета. Чтобы избежать потерь в информации по часто измеряемым и учитываемым параметрам, можно использовать наблюдения с разной длительностью кванта. Тогда в системе рекурсивных моделей выходные показатели моделей с малой единицей наблюдения выступают как независимые переменные моделей с более продолжительной единицей наблюдения.

Выбор временных границ совокупности и числа квантов, включаемых в нее, определяется организацией представительной временной выборки — достаточной для построения адекватной модели. Протяженность периода связана с уровнем объекта и величиной кванта наблюдения, с назначением модели, устойчивостью техники и технологии производства, со стабильностью внешних экономических связей и факторов. Чем выше уровень объекта и короче квант, тем меньше календарная длительность периода анализа; существенные изменения в способе

производства ограничивают продолжительность периода исследования.

Пространственные границы совокупности могут задаваться либо перечнем объектов, входящих в исследуемую совокупность, либо набором признаков, общих для всех объектов совокупности. При выборе пространственных границ учитываются, с одной стороны, цель исследования, предъявляемые к будущей модели требования, с другой — требования логики статистического моделирования (определенная однородность, достаточный объем совокупности). Если пространственные границы заданы перечнем объектов, то они со временем должны изменяться (ввод в строй новых предприятий). Пространственные границы, заданные в виде набора признаков, более устойчивы во времени. Набор признаков, используемых для ограничения совокупности, зависит от уровня объектов моделирования. Чем выше уровень, тем меньше набор ограничивающих признаков. При построении пространственных границ определенное значение имеет административная принадлежность объектов, так как результаты моделирования затем практически должны использоваться в рамках определенной административной системы.

2.3. Априорный теоретический анализ предполагает определение набора производственных факторов и экономических показателей, описывающих процесс функционирования исследуемых объектов, и изучение структурных связей между ними. Выбор моделируемых показателей (одного или нескольких) сравнительно прост: они задаются самой постановкой задачи, формулировкой ее цели, во многих случаях необходимым набором показателей определяется самим назначением модели. Например, если статистическая модель объекта должна входить составной частью в двухуровневую модель «предприятие — отрасль», то набор моделируемых показателей определяется выбранным типом отраслевой модели (модели верхнего уровня).

Процесс формирования набора производственных факторов носит характер последовательно уточняемой гипотезы. Основанием для выдвижения первоначальной гипотезы служат теория (экономическая и моделируемого производства), результаты предшествующих исследований, профессиональный опыт специалистов отрасли. Целесообразно сначала установить перечень укрупненных характеристик функционирования исследуемых объектов, а затем

детализировать их. В дальнейшем исследовании набор характеристик сужается: отдельные факторы не поддаются измерению; в ходе расчетов выясняется существенность новых факторов; сказываются соображения экономической целесообразности, требования и ограничения, накладываемые математическим аппаратом и применяемой вычислительной техникой.

Перечень укрупненных факторов для моделирования показателей мартеновского производства следующий:

1. Объем, структура и качество основных фондов цеха (технические параметры цеха и печей, возрастной состав и степень освоенности фондов).
2. Количество, структура и качество сырьевых материалов (основных и добавочных).
3. Виды и качество топлива.
4. Характеристики трудовых ресурсов.
5. Технологические способы выплавки и разливы стали.
6. Параметры теплотехнического режима эксплуатации печей.
7. Природно-географические условия работы.

В дальнейшем эти комплексные характеристики разветвлялись в наборы более детализированных факторов. Например, характеристика сырьевых материалов включала соотношение чугуна и лома, химический состав чугуна, качество и структуру лома, виды и качество добавочных материалов. Технологию плавки и разливы представляли вид процесса (скрап-процесс, скрап-рудный и дуплекс), применение интенсификаторов (кислород, сжатый воздух), способы разливы (стационарные канавы, изложницы на тележках, непрерывная разлива, разлива сверху и сифоном) и т. д. Развернутый набор факторов служил исходной базой для последующего уточнения гипотезы.

2.4. Выявлению структуры причинно-следственных связей нередко препятствует неизученность логики производственного процесса. Однако практически почти всегда могут быть высказаны некоторые профессиональные соображения о характере и направлении связей. Связи набора признаков, описывающих процесс функционирования реального объекта, всегда имеют определенную структуру и ориентацию. Это позволяет формально представить структуру связей, для чего можно использовать аппарат теории графов (Берк, 1962). При таком подходе каждой изуча-

емой системе переменных ставится в соответствие определенная логическая структура, формализуемая некоторым графом (матрицей смежности). При помощи операций по преобразованию графов (матрицы смежности) можно уменьшить размерность набора переменных, не искажая структуру причинно-следственных связей оставшихся переменных, т. е. можно выделить необходимую часть структуры так, чтобы связи, опосредованные исключаемыми характеристиками (переменными), сохранились в новой структуре. В ходе анализа матрицы смежности выделяются слои переменных по характеру влияния (структура входа, структура выхода, промежуточные слои) и выявляются различные типы связей. Это, в свою очередь, позволяет более обоснованно подходить к формированию совокупности, к выбору метода моделирования и типа модели (классические регрессионные модели, рекурсивные и структурные уравнения и т. п.). Метод анализа связей переменных во многом зависит от того, к какому типу структур они принадлежат.

Таким образом, построение, преобразование и упорядочение матрицы смежности означает уменьшение случайной хаотичности и внесение стройности в постановку задачи, отвечающей объективным закономерностям исследуемой системы.

§ 2. Организация экспертного опроса и анализ его результатов

2.5. При решении задач априорного анализа целесообразно использовать накопленный опыт специалистов с помощью методов экспертных оценок. Наиболее удобная форма его реализации — анкетный опрос.

Анкета обычно содержит три рода сведений: основные — ответ эксперта о порядке расположения (информативности) признаков, вспомогательные — характеристики самого опрашиваемого (должность, профессиональный стаж, образование, тип предприятия, которое он представляет, и т. п.) и различного рода справочные реквизиты и пояснения (название анкеты, обращение, пояснения о порядке заполнения и т. д.).

По характеру организации опроса следует выделить три его формы. Первая: каждому опрашиваемому специалисту предлагается назвать неограниченное количе-

ство признаков, влияющих на исследуемый показатель, и оценить степень их влияния. Вторая: исследователь заранее составляет перечень признаков, а задача эксперта сводится к оценке степени влияния каждого из названных признаков на моделируемый показатель и к их ранжированию.

В третьей, более сложной форме организации опроса последовательно используются оба названных способа анкетирования. Сначала коллективу экспертов предлагается свободно назвать признаки и дать им оценку, затем на основе анализа результатов первого тура опроса формируется анкета с заданным перечнем признаков для оценки их значимости на втором туре.

На практике наиболее широко используется второй из указанных способов, так как он экономит время опрашиваемых, дает возможность исследователю более строго классифицировать ответы и сравнительно прост по реализации. Все дальнейшее изложение ведется применительно к этому способу организации опроса.

2.6. При построении перечня признаков необходимо обеспечить достаточную представительность предлагаемого набора, не упустить каких-либо существенных признаков и дать четкие недвусмысленные названия каждого из них.

Первоначальный набор оцениваемых признаков должен описывать различные стороны процесса производства. Если совокупность признаков отражает лишь одну сторону процесса, определить «объемную» конструкцию не смогут даже высококвалифицированные эксперты.

Довольно сложно выбрать «хорошие», понятные названия признаков. Выражаемые словесно, оцениваемые признаки полисемантичны (т. е. разные люди могут вкладывать в них различный смысл). Поэтому надо стремиться к использованию установившихся, общепринятых в данной специфической области определений и понятий, а также к проведению предварительных, пробных опросов, по результатам которых уточняется первоначальный перечень.

2.7. Ранжировка признаков проводится экспертом на основе выбранной составителем анкеты шкалы оценки влияния признаков на моделируемый показатель (вектор показателей).

Самой простой шкалой может быть двухбалльная: влияют на показатель и не влияют. Однако влияние признака может изменяться от весьма малой до значительной вели-

чины. В этих случаях применение двухбалльной шкалы оказывается весьма грубым приемом и ведет к значительной потере информации. Поэтому на практике часто используется многобалльная шкала оценок, когда вводится ряд промежуточных состояний. Однако субъективные особенности человеческого мышления накладывают на количество оцениваемых мест ограничения как снизу (не менее 5—6), так и сверху (не более 12—15).

Чтобы избежать введения чрезмерно большого числа градаций шкалы оценок и присвоения многим признакам одного места, иногда используют двухступенчатые перархические анкеты. Признаки объединяются в группы, которые эксперт ранжирует, а затем определяет в них место отдельных признаков. При статистической обработке таких анкет возникают определенные трудности, если по ходу исследования необходимо построить единую систему расположения всех признаков.

В этом случае приходится прибегать к некоторым искусственным способам формирования общих оценок роли признаков. К таким способам можно отнести:

последовательное расположение групп и отдельных факторов внутри них по их значимости;

ранжировку факторов по сумме баллов — в основе ее лежит величина, получаемая при сложении балла группы с баллом, присвоенным фактору внутри группы;

введение «сравнимых» шкал для меж- и внутригрупповых оценок. Здесь используется следующий прием: суммируются оценки факторов внутри каждой группы и на эту сумму делится балл группы. Полученная величина представляет вес единицы внутригрупповой шкалы. Умножая на этот коэффициент оценки каждого фактора, получаем скорректированные оценки, по которым выполняется ранжировка.

2.8. Статистический анализ материалов анкетного опроса предполагает группировку, агрегирование признаков; оценку степени согласованности ответов экспертов по каждому признаку в отдельности и в целом по всему набору; выделение групп экспертов с «близким» мнением о порядке признаков в случае существенных расхождений в ответах; выявление причин разброса мнений, определение влияния характеристик экспертов на содержание ответов; оценивание качества экспертных оценок и компетентности экспертов; формирование группового решения.

Группировка признаков необходима потому, что, во-первых, отдельные признаки эксперты не различают по информативности; во-вторых, с устранением случайных ошибок измерения роли отдельных признаков оценка значимости влияния группового фактора получается более точной; в-третьих, эксперты по-разному могут истолковывать названия признаков.

Ответы экспертов представляют собой балльные оценки каждого признака. При обработке подобной информации в первую очередь надо принять определенную шкалу измерения. Это означает известную идеализацию балльных оценок: в форме количественного признака — при принятии метрической шкалы и классификационного — для неметрических шкал. В последнем случае имеет место значительная потеря информации. Из неметрических шкал наиболее адекватна порядковая.

В основу группировки признаков кладется матрица их взаимосвязи, формализующая понятие близости между признаками. В качестве мер связи в зависимости от характера идеализации балльных оценок могут выступать коэффициенты корреляции (сильное предположение о метрической шкале), взаимной сопряженности Пирсона и Чупрова (упрощающее предположение о качественном характере оценки), ранговой корреляции (предположение о порядковой шкале) и т. д. (Юл, Кендалл, 1960; Устюжанинов, 1969; Миркин, Черный, 1969; и др.). Полезно использовать разные (две — три) меры связи и оценить степень устойчивости полученных группировок.

Группы признаков выделяются каким-либо алгоритмом таксономии, рассчитанным на работу с матрицей связи. При большом числе признаков весьма удобен для решения указанной задачи метод экстремальной группировки (Браверман, 1970), при использовании в качестве мер близости коэффициентов корреляции — методы факторного анализа (Лоули, Максвелл, 1967).

Для анализа согласованности ответов экспертов по отдельным признакам исходные данные опроса удобно представить в виде набора вариационных рядов ответов по каждому вопросу. Обозначим члены ряда через f_{ij} (f_{ij} — число ответов о присвоении i -му признаку j -го места). Оценка степени согласованности ответов дополняет оценку уровня вариации. Если обозначить некоторый показатель вариации признака i через μ_i , то мера согласованно-

сти по этому признаку будет $1 - \mu_i$. Для расчета величины μ_i можно применить общеизвестные статистические приемы оценки вариации в рядах распределения, а именно измерение: области, содержащей основную часть ответов; отклонений переменных от центрального значения (средней, медианы); степени однородности качественных переменных.

Природе балльных оценок наиболее соответствуют первый и третий подходы. При первом в качестве меры согласованности ответов может быть, например, использован коэффициент, равный отношению числа ответов, приходящихся на медианное и два соседних с ним места, к общему числу ответов. Значительно большую информацию обеспечивает третий подход, когда вычисляются коэффициенты вариации качественных переменных. Расчет таких коэффициентов основан на сравнении фактического числа различных пар событий с максимально возможным числом пар.

Подобная мера вариации качественных признаков предложена Д. Мюллером и К. Шусслером (1968) и представляет аналог дисперсии

$$\mu_i = k/(k-1) \times \left[\left(\sum_j f_{ij} \right)^2 - \sum_j f_{ij}^2 \right] : \left(\sum_j f_{ij} \right)^2,$$

где μ_i — коэффициент качественной вариации i -го признака; k — число градаций (число мест) i -го признака. Величина μ_i меняется в пределах $0 \leq \mu_i \leq 1$.

Анализ характера распределения ответов по отдельным признакам позволяет выдвинуть предварительную гипотезу о причинах разброса мнений. Здесь возможны следующие ситуации.

Распределение имеет одну вершину, степень согласованности ответов достаточно высока. Можно полагать, что разброс мнений вызван случайными причинами.

Просматриваются несколько четко выраженных вершин распределения с провалами между ними. Есть основания предположить наличие нескольких групп экспертов с разным характером мнений, что важно учитывать при выделении однородных групп экспертов на следующих этапах статистической обработки анкет.

Ответы сравнительно равномерно распределяются по всей шкале оценок, нет четко выраженных вершин, степень согласованности мала. Можно предположить, что ли-

бо в анкете дана нечеткая формулировка названия признака, либо введено слишком много градаций шкалы, либо эксперты некомпетентны в этом вопросе.

В чистом виде перечисленные ситуации встречаются редко, чаще имеют место их комбинации.

В качестве иллюстративного примера приведем результаты оценки вариации ответов, которые дали 87 специалистов маргеновского производства на вопрос о силе влияния групп производственных факторов на себестоимость стали (табл. 2.1). Их анализ позволяет заключить, что разброс полученных ответов довольно значителен, а хорошая согласованность мнений из 9 групповых признаков таблицы получена по 1, 2 и 9-му («техническая характеристика цехов», «объем производства и сортамент продукции», «географическое положение»).

Наибольшее единообразие в ответах наблюдается по группам признаков, занимающих крайние места. По мере приближения к средним местам различия в мнениях растут (мера согласованности от 0,31 для 1-го места снижается до 0,04 для 5-го и затем возрастает до 0,16 для 9-го).

По причинам разброса мнений экспертов к первой ситуации можно отнести распределение ответов по 1, 2 и 9-му признакам, ко второй — распределение по 4, 7-му признакам (просматриваются две вершины), к третьей — ответы по 5-му признаку и отчасти по 6-му.

Необходимость в выделении однородных групп экспертов связана с конечной целью экспертного опроса — формированием группового решения на основе совокупности индивидуальных оценок. Традиционные способы принятия группового решения (принцип большинства средних, центральных оценок) применимы только в тех случаях, когда коллектив экспертов однороден в смысле «поведения» (характера ответов). Если коллектив неоднороден, то средние, центральные оценки теряют содержательный смысл, они могут оказаться в определенном смысле «хуже» индивидуальных или групповых, на основе которых получены. В таком случае либо отдают предпочтение точке зрения, которая представляется более соответствующей теоретическим представлениям, либо разбивают совокупность объектов на части и приписывают каждой части отдельную шкалу.

Группировка экспертов по характеру их ответов предполагает введение меры сходства каждой пары экспертов. При использовании любой из возможных мер сходства в результате расчетов явно или неявно получаем квадратную матрицу мер близости (расстояний) между экспертами по характеру их ответов. Следующий этап заключается в ис-

Таблица 2.1

Оценка вариации ответов экспертов о влиянии групп производственных факторов на себестоимость стали

Групповой признак	Занимаемое место									Итого	Мера согласованности на трех местах, включая среднее, %				
	1	2	3	4	5	6	7	8	9			10	11	12	
Техническая характеристика цехов.	50	45	40	4	3	2	2	2	1	—	—	—	87	84,5	0,31
Объем производства и сортамент продукции	21	22	26	6	3	5	1	1	1	—	—	1	87	79,3	0,15
Технология плавки и разливки.	7	44	49	23	10	8	4	1	1	—	—	—	87	64,4	0,08
Характеристика сырых материалов.	40	16	41	20	14	9	2	2	1	—	—	2	87	54,0	0,07
Технико-экономические показатели.	43	7	43	4	16	40	8	6	8	1	1	—	87	37,9	0,04
Технологические параметры.	3	10	40	41	44	21	9	5	1	—	—	3	87	52,9	0,08
Возрастная структура фондов.	4	—	—	7	40	23	16	20	7	—	—	—	87	56,3	0,05
Характеристика кадров цеха.	—	2	1	7	15	3	18	17	8	3	1	2	87	55,2	0,07
Географическое положение.	4	2	—	5	3	1	9	11	38	12	—	5	87	66,7	0,16

следовании этой матрицы и — при необходимости — в разбиении ее на однородные группы одним из алгоритмов таксономии.

С этой целью выделяется устойчивое ядро, включающее основную часть экспертов, а постепенно отделяющиеся от него (основного таксона) эксперты не образуют устойчивых таксонов, по численности сравнимых с основной группой. (Вполне понятно, что эксперты, отделяющиеся на первых шагах, обладают наиболее оригинальным, резко отличным от большинства экспертов мнением.) В качестве эталонной группы следует выбрать ядро экспертов, сформировавшееся на четвертом шаге процесса, после которого резко возрастает число единичных таксонов.

Для выделения однородных групп экспертов по материалам опроса специалистов мартеновского производства использовали алгоритм и программу таксономии «Форель-1» (Елкина, Загоруйко, 1966). Подлежащая таксономии совокупность состояла из 18-мерных векторов ответов (x_1, x_2, \dots, x_{18}), полученных от 87 экспертов. Результаты разбиения на отдельных шагах приведены в табл. 2.2.

Таблица 2.2

Результаты выделения однородных групп экспертов по мартеновскому производству

Характеристика таксономии	Номер шага					
	0	1	2	3	4	5
Радиус гиперсферы	5,93	5,34	4,75	4,15	3,56	2,97
Число таксонов, всего	1	4	10	19	30	48
В том числе:						
единичных	—	3	7	13	24	38
включающих 2—6 экспертов	—	—	2	5	5	9
Число экспертов в основном таксоне	87	84	73	60	52	23

Выявление зависимости ответов экспертов от их специальных характеристик целесообразно начинать с оценки связи двух разбиений совокупности экспертов: по характеру ответов и по объективным характеристикам (профессия, занимаемая должность, тип представляемого предприятия, стаж работы, образование, возраст и т. п.). Если связь между этими разбиениями существенна, то на втором этапе определяются наиболее важные по влиянию на ответы характеристики экспертов.

Для оценки роли отдельных характеристик экспертов в возникновении неоднородности ответов можно использовать разные способы. При небольшом числе характеристик общая совокупность анкет разбивается на группы по каждой из этих характеристик. Затем для набора специальных групп, образованных разбиением совокупности анкет по определенной характеристике экспертов (тип предприятия, специальность и т. п.), рассчитываются коэффициенты общей согласованности мнений (типа коэффициентов корреляции или их аналогов). Сравнительный анализ величины коэффициентов конкордации по группам при разбиении совокупности по разным характеристикам позволяет определить относительную важность отдельных характеристик экспертов. Характеристика, которая обеспечивает самую большую внутригрупповую согласованность ответов, будет наиболее важной.

Другой способ состоит в том, что многомерная классификация экспертов по ответам считается заданной и определяются характеристики экспертов, наиболее информативные для этого разбиения. Веса значимости отдельных характеристик экспертов определяются по одному из алгоритмов минимизации описания, в частности по алгоритму СПА (Лбов, 1965).

Например, существенные отличия эталонной группы специалистов мартеновского производства от всей совокупности выявились по двум характеристикам — принадлежности к отдельным профессиональным группам и стажу работы в металлургии. Составим профессиональную структуру эталонной группы и всего коллектива экспертов (в %):

	Вся совокупность	Эталонная группа
Работники экономических служб	23,8	32,0
Работники технологических служб	25,0	30,0
Исследователи	25,0	18,0
Главные специалисты	26,2	20,0

Исследования показали, что наиболее едины в своих мнениях работники предприятий, непосредственно занимающиеся анализом и планированием производства. Что касается стажа работы, то эталонная группа отличается повышенной долей лиц с очень большим стажем (свыше 20 лет) за счет уменьшения доли лиц со средним стажем (10—20 лет).

§ 3. Сбор и подготовка исходной информации

Результаты экономико-статистического исследования целиком обусловлены качеством исходных данных — их полнотой и достоверностью. Существующая обширная производственно-экономическая статистика предприятий представляет собой только потенциальную научную информацию. Чтобы стать реальной, она должна быть проанализирована на соответствие целям и теоретическим посылкам исследования. Кроме того, следует выбрать и рассчитать необходимые операционные показатели, обеспечить их возможную степень сопоставимости, оценить (хотя бы на качественном уровне) точность и достоверность информации.

2.9. При формировании исходной информации на базе отчетных данных прежде всего выбираются или конструируются для каждого фактора рабочего набора соответствующие измерители, которые будут выступать затем как переменные будущей модели.

Большинству производственных факторов присуща сложная внутренняя структура, которая может быть описана комплексом измерителей. Например, сортамент продукции представляет собой совокупность долей отдельных видов продукции; простой агрегатив — совокупность времени бездействия по различным техническим и организационным причинам; для описания мощности пласта в угольной промышленности используются три измерителя: горно-геологическая, средневывнимаемая и среднединамическая мощность и т. п. Значение фактора может быть измерено в натуральных или относительных единицах. В свою очередь, натуральные единицы могут варьировать в зависимости от физического содержания (например, расход топлива — в физических и условных тоннах), от исторически сложившихся особенностей меры (метр — дюйм, кВт·ч — л. с.), от базы, принятой за основу расчета, и т. д.

Набор выбранных измерителей должен обеспечивать: сопоставимость измерителей в пространстве и времени; достаточно полную характеристику описываемого ими фактора;

общность для всех единиц совокупности; предпочтительность количественных оценок перед качественными, натуральных перед стоимостными;

учет специфических задач исследования (например, разделение объективных и субъективных влияний).

Использование методов математической статистики предполагает, что по каждому объекту анализируемой совокупности информация представлена одинаково. Эти методы нейтральны по отношению к тому, насколько внешняя одинаковость соответствует единству экономического содержания. Данные действующего производственно-технического учета часто не обеспечивают соблюдения сопоставимости из-за различий в уровне агрегирования показателей (отчетных сведений); в структуре единиц совокупности; в способах учета показателей во времени; в периодичности учета отдельных показателей (разная длительность рабочего времени в календарных отрезках времени — неделя, месяц, квартал, год); в принципах и процедурах, лежащих в основе исчисления производных показателей; в ценах и других внешних экономических условиях функционирования объектов; наконец, в масштабах производства.

2.10. Способ устранения несоответствия, возникающего из-за разной степени агрегированности показателей, зависит от того, на какой стадии измерения и учета происходит агрегирование факторов. Для факторов, по которым первичный учет на предприятиях ведется дифференцированно, а агрегирование происходит лишь при составлении отчетности, он состоит в обращении к первичным документам. (При систематическом ведении подобных исследований целесообразно отчетность оформлять в соответствии с необходимой степенью дифференциации факторов.)

Другие факторы, например некоторые характеристики работы мартеновских печей, измеряются и учитываются сразу на цеховом уровне (расход топлива и сырых материалов). Чтобы исправить это положение, формируется такая подсовокупность объектов, по которой применение усредненных сведений было бы оправданным. Для устранения излишней агрегированности можно также условно распределить агрегированный показатель между объектами более низкого масштаба, пропорционально значениям какого-либо дифференцированного показателя — емкости печей, длительности плавки и т. п. Вследствие искусственности этого приема сфера его применения ограничена.

Наконец, маловажные факторы, представленные в отчетности среднезаводскими данными, следует исключить

из рассмотрения (они могут вносить в модели искажения).

Проблема различной агрегированности факторов представляется наиболее острой для промышленных объектов низкого уровня (технологические агрегаты). По мере повышения уровня объектов (цехи, предприятия) острота проблемы снижается, хотя и для предприятий в условиях образования фирм, объединений, комбинатов с общей отчетностью эта проблема также может возникать.

2.11. Различия в структуре единиц совокупности наибольшие трудности вызывают при исследовании пространственной выборки, где приходится оперировать со многими объектами разного масштаба и структуры. Однако несопоставимость структуры, хотя и реже, может возникнуть и при исследовании работы отдельного объекта во времени (реконструкция, измерение административной подчиненности и т. п.). Чем выше уровень объекта моделирования, тем вероятнее структурные различия (эта вероятность меньше у технологических агрегатов и значительно больше у предприятий и комбинатов). Обеспечить сопоставимость структуры исследуемых объектов — значит четко определить объект наблюдения. В некоторых случаях здесь необходим либо переход к моделированию объектов более низкого уровня с сопоставимой структурой, либо составление системы структурных уравнений.

Существует два способа учета показателей во времени: оценка состояния объекта на определенный момент времени и учет результата деятельности объекта за определенный период.

Описывая процесс функционирования объекта, необходимо по возможности стремиться к тому, чтобы все переменные модели были измерены одинаково и в понятиях результата, расхода. Если это по каким-либо обстоятельствам невозможно, надо по крайней мере попытаться уменьшить величину ошибки, возникающей вследствие различия в способах измерения и из-за использования моментных оценок. Замена показателей расхода (результата) показателями наличия (состояния) допустима лишь при их сильной взаимосвязи, например между численностью персонала и отработанным временем. При этом не следует забывать о содержательном различии показателей.

Если периодичность учета отдельных переменных разная и предполагается строить модель из одного уравнения, то приходится ориентироваться на переменные с наиболь-

шей величиной интервалов замера. По переменным с более частыми замерами при этом возможны значительные потери информации. Поэтому рационально строить систему рекурсивных уравнений, в которой начальное уравнение содержит переменные, имеющие наименьший интервал между измерениями, а выходной показатель этой модели выступает как аргумент в моделях следующего уровня агрегирования с более длительными интервалами между замерами, и т. д.

2.12. При использовании стоимостных показателей для получения надежных оценок важно учитывать различия в ценах. С этой целью чаще всего показатели пересчитываются в единых ценах. Включение ценностного фактора в исследование зависит от постановки задачи. Если задача рассматривается широко — в отраслевом масштабе, с учетом межзаводских связей, тогда включение цен может быть необходимым. Если же изучается влияние чисто производственных факторов, то представление себестоимости в действующих ценах приведет к искажению результатов и необходим пересчет стоимостных показателей для всех объектов в единых ценах. В качестве базовых могут служить цены одного из предприятий.

2.13. Источников информации о функционировании объекта в основном два: непосредственные наблюдения, регистрация интересующих исследователя сведений и ранее накопленная информация, которая содержится в первичной учетной и технической документации, в диаграммах самопишущих приборов и т. п.

Соответственно этим источникам различаются два метода сбора информации — активный и пассивный эксперимент. В первом случае планирование и управление процессом осуществляется самим исследователем, он выбирает некоторую оптимальную стратегию для управления экспериментом. Процесс исследования обычно разбивается на отдельные этапы. После каждого этапа исследователь получает новую информацию, позволяющую ему изменить стратегию управления.

Во втором случае сбор информации осуществляется после свершения события. Единственно возможная форма контроля со стороны исследователя — применение какого-либо метода целенаправленного отбора из имеющихся данных такой информации, которая отражает независимый от исследователя ход процесса.

Активный эксперимент позволяет задавать и контролировать необходимый набор переменных, их уровни и комбинацию. Однако при моделировании работы крупных промышленных объектов он технически не всегда может быть организован, и единственно возможным способом сбора информации оказывается пассивный эксперимент.

2.14. Надежность исходной информации зависит от ошибок, вероятных на каждом этапе подготовки данных от их появления до использования в качестве исходной информации исследования. Можно выделить четыре таких этапа: измерение, сбор и первичную обработку, агрегирование, подготовку и проведение математических операций.

На первом этапе ошибки вызываются различного рода неполадками в работе измерительных приборов. Так как происходит постоянное совершенствование приборов, то эти ошибки уменьшаются.

Этап сбора и первичной обработки — источник стихийных ошибок, связанных с фиксацией данных человеком. Практика показывает, что подобные ошибки весьма многочисленны и, в связи со случайной природой возникновения, их нелегко выявить и устранить. Важно каким-то образом приблизительно оценить вероятный уровень таких ошибок и учитывать его при анализе результатов экономико-статистического исследования. В сущности, это связано с решением проблемы необходимой точности модели.

На этапе агрегирования первичные данные сводятся в комплексы по временному или пространственному принципу, например подсчитываются среднегодовые величины, группируются продукты по технологии изготовления или назначения, показатели работы производственных участков объединяются в показатели административного подразделения и т. п. Здесь ошибки обуславливаются как нечеткостью в определениях временного интервала и структуры пространственного объединения, так и трудностями в отнесении самих данных к тому или иному комплексу. Ошибки подобного рода, конечно, ведут к очень серьезным искажениям в статистической информации — получению несравнимых данных, двойному счету и т. п. Для их устранения необходима более углубленная разработка теоретических вопросов (определение и классификация отраслевой экономики) и создание подробно информированных в этом отношении статистических служб.

Ошибки этапа математической обработки данных можно подразделить на две группы: ошибки подготовительных операций (кодирование, перфорирование) и связанные с надежностью работы ЭВМ (сбой машины, неполадки в печатающем устройстве и др.). Устранить ошибки второго рода можно путем дублирования результатов, но тогда значительно увеличиваются затраты на исследование. Более экономичный путь состоит в разработке и использовании алгоритмов, которые предусматривали бы контроль надежности получаемых результатов.

§ 4. Предварительный статистический анализ и преобразование исходной информации

2.15. Основные задачи первичного статистического анализа моделируемых показателей и характеристик производственного процесса заключаются в изучении их уровней, характера вариации и статистических взаимосвязей, а также в выделении взаимосвязанных комплексов и формировании информативных подпространств.

Методика решения этих задач с помощью корреляционно-регрессионного анализа применительно к количественным признакам описана во многих работах, например Я. И. Лукомского (1958), Н. К. Дружинина (1974), М. Езекеела и К. Фокса (1966) и др.

В экономико-статистических исследованиях могут применяться три типа признаков — классификационные, ранговые и количественные.

Классификационные признаки определяют неупорядоченное разбиение множества объектов. При исследовании можно лишь отмечать наличие или отсутствие данной градации признака у единиц совокупности и подсчитывать число единиц, обладающих или не обладающих этой градацией.

Ранговые признаки порождают упорядоченное разбиение совокупности на классы. Допускаются операции «равенство — неравенство» и «больше — меньше».

Признак называется количественным, если его значения — числа и он может быть измерен по каждой единице совокупности. Допускаются все арифметические операции. При необходимости количественные признаки могут быть перекодированы в качественные (ранговые и классификационные), естественно, с определенной потерей

информации. Для перевода в ранговые следует разбить количественный признак на интервалы и присвоить им ранги в соответствии с их величиной. Чтобы уменьшить потери информации при перекодировке, иногда полезно разбивать количественные признаки на неравные интервалы соответственно концентрации их значений.

2.16. Для соизмерения признаков различной природы разрабатывается формальный аппарат, пригодный для измерения связи между признаками различных типов (Миркин, Черный, 1972). Универсальность введенной указанными авторами меры нуждается в практическом опробовании.

Из приемлемых в настоящее время путей соизмерения различных признаков следует отметить два: перекодировку количественных признаков в качественные (классификационные, ранговые) с использованием коэффициентов связи качественных признаков и формирование количественного эквивалента качественных признаков на базе теории распознавания образов.

При использовании первого способа составления таблицы взаимной сопряженности каждой пары признаков и рассчитывается критерий χ^2 , характеризующий отклонения наблюдаемых частот в клетках таблицы от ожидаемых, которые оцениваются исходя из предположения о независимости признаков (Митропольский, 1974; Юл, Кендалл, 1960; Чупров, 1962).

В основе формирования количественных эквивалентов лежит предположение, что качественные различия находят отражение в измеряемых характеристиках процесса и косвенно могут быть ими представлены. Процедура поиска количественного эквивалента включает четыре операции:

1. Разбиение совокупности на группы по выделенным существенным качественным признакам (или признакам-представителям) — получение обучающей выборки.

2. Формирование описания полученного разбиения в пространстве количественных признаков — обучение распознаванию выделенных классов в другом пространстве. Результаты обучения позволяют выяснить, можно ли на основе имеющегося набора количественных признаков построить достаточно точный эквивалент изучаемого качественного признака (их комбинации). Точность эквивалента оценивается, например, числом ошибок распознавания

градаций качественных признаков объекта по значению его количественных признаков. Если построение эквивалента невозможно, ищутся способы совместного учета качественных и количественных признаков.

3. Выбор эффективной подсистемы количественных признаков меньшей размерности, описывающих классы с достаточной точностью.

4. Конструирование единого эквивалента из количественных признаков, вошедших в эффективную подсистему.

Например, при исследовании мартеновского производства для подбора общего количественного эквивалента признаков «тип процесса» и «способ выплавки» на первом этапе применили комбинационную группировку. В результате были сформулированы три исходных образа (класса): I — цехи, работающие на твердом чугуне без применения кислорода; II — цехи, работающие на жидком чугуне без применения кислорода; III — цехи, работающие на жидком чугуне с применением кислорода.

На втором этапе исходный материал обучения по каждому образу задавался в виде матрицы, элементы которой представляли значения количественных признаков для цехов, принадлежащих данному образу. Полученное по всему набору количественных признаков описание образов (классов) и есть количественный эквивалент. О его высокой точности свидетельствует хорошая аппроксимация заменяемых качественных признаков — на экзамене при распознавании номера образа для 65 цехов сделана только одна ошибка.

На третьем этапе была выбрана эффективная подсистема из шести признаков (доля чугуна в садке; удельный расход кислорода; расход окислителей; доля стали, выплавляемой с кислородом; удельный расход топлива; срок действия цеха) с весами соответственно 0,323; 0,260; 0,214; 0,090; 0,023.

2.17. Снижение размерности признакового пространства в предмодельный период обусловлено рядом обстоятельств. Во-первых, необходимо устранить явление мультиколлинеарности, заключающейся в том, что исходный набор включает пары или группы признаков, между которыми имеется функциональная или весьма тесная линейная корреляционная зависимость.

Во-вторых, должно учитываться максимально возможное для данных условий число переменных модели. При ограниченном числе объектов переменных в модели должно быть в несколько раз меньше, чем единиц наблюдения.

В-третьих, сказываются требования и ограничения, накладываемые применяемой техникой и используемыми алгоритмами обработки информации; соображения снижения

трудоемкости исследования, удобства построения и последующей интерпретации модели. Конечная цель минимизации описания более глубока, нежели просто коротко представить большой массив исходных данных. При представлении информации может быть вскрыта объективная закономерность, существующая в структуре признакового пространства и позволяющая провести это сокращение.

Максимальный набор производственных факторов на этапе подготовки информации для моделирования может быть сокращен на основе содержательных представлений об изучаемом производственном процессе; в связи с проверкой обеспеченности и качества информации; с помощью формальных методов статистического анализа. О масштабах возможного эффекта можно судить по результатам исследования мартеповского производства: отсев малозначимых и с низким качеством информации факторов позволил сократить их набор с 87 до 47. При сокращении формальными методами выделен для стадии моделирования набор из 21 производственного фактора. Однако это не должно наводить на мысль, что формирование максимального набора излишне. Именно наличие максимального набора и позволяет выделить взаимосвязанную систему наиболее информативных признаков благодаря возможности перебора огромного числа их сочетаний в подсистемах.

Процесс выбора существенных факторов не заканчивается на этапе подготовки информации, а продолжается и при построении модели. Но идеи, лежащие в основе выбора информативных факторов, на этих стадиях процесса иные. При подготовке информации сужение набора переменных осуществляется исходя из внутренних свойств связей факторов производства. Поиск информативной подсистемы переменных в ходе построения модели кроме взаимосвязи факторов учитывает силу их влияния на функцию — моделируемый экономический показатель. Однако резкой грани между двумя стадиями выбора существенных признаков нет. Они взаимосвязаны, часты возвраты к предшествующей стадии по результатам последующей.

2.18. Сокращать признаковое пространство можно на основе либо исходной матрицы с помощью общих алгоритмов таксономии¹, либо анализа квадратных матриц пар-

¹ Обзор алгоритмов можно найти в работах С. А. Айвазяна, З. И. Бежаевой, О. В. Старовой (1974); А. А. Дорофеева (1971).

ной связи (расстояний) признаков между собой. Ориентировка на использование матриц связи обусловлена тем, что на их основе могут решаться задачи анализа структуры связей признаков, существенно упрощаются некоторые алгоритмы анализа, сами показатели связи могут использоваться для дальнейших расчетов параметров регрессионной модели.

Показателями связи выступают коэффициенты корреляции количественных признаков, корреляции ранговых признаков, взаимной сопряженности качественных признаков, различного рода энтропийные меры связи, мера близости двух разбиений и др.

В случае симметричных матриц связи группировка признаков сводится к перестройке матрицы таким образом, чтобы вдоль главной диагонали выделялись квадратные блоки с большими значениями показателей силы связи, а элементы, расположенные вне блоков, были относительно малы. Такая перестановка строк и столбцов матрицы связи признаков получила в литературе название «диагонализации» (Браверман, Дорофеев, Лумельский, Мучник, 1971). Ее результат — такое объединение признаков, что входящие в одну группу обладают большей силой связи между собой, чем принадлежащие к разным группам. Общепризнанной классификации алгоритмов (диагонализации) нет, условно их можно подразделить на эвристические и строгие вариационные, основанные на поиске экстремума некоторого функционала, характеризующего качество разбиения.

Сложность вариационного подхода состоит в том, что некоторые простые и кажущиеся естественными критерии часто оказываются неинтересными с содержательной точки зрения. Кроме того, разные функционалы приводят к неодинаковым результатам диагонализации матрицы и к различной содержательной интерпретации понятия «сильной» внутрigrупповой связи. Это обстоятельство обуславливает широкое использование эвристических алгоритмов для матриц небольшой и средней размерности².

Приведем пример сужения набора первичных факторов (35) при моделировании себестоимости стали.

Набор сокращался с использованием метода факторного

² См., например, работы Э. М. Бравермана (1970), А. В. Беккера, М. Л. Лукацкий (Вопросы экономико-статистического моделирования. Новосибирск, 1970, с. 273—284) и других авторов.

анализа, а именно центроидного метода (Лоули, Максвелл, 1967), позволяющего разбить признаки на заданное число групп и для каждого признака определить его коэффициент корреляции с условным средним признаком группы. Степень взаимосвязанности признаков в группе и изолированности групп друг от друга определяется задаваемым числом групп. Проведено два экстремальных разбиения признаков на 6 и 12 групп.

При разбиении на 6 групп состав последних и теснота связи (r) со средними признаками групп были следующими:

Фактор		r	Фактор		r
1-я группа	x ₇	0,727	4-я группа	x ₅	0,714
	x ₁₂	0,576		x ₉	0,760
	x ₂₅	0,483		x ₁₄	0,656
2-я группа	x ₂₉	0,589	x ₁₅	0,690	
	x ₈	0,576	x ₂₁	0,886	
	x ₁₆	0,545	x ₂₂	0,952	
	x ₂₆	0,460	x ₂₇	0,819	
	x ₃₃	0,626	5-я группа	x ₁₀	0,874
x ₃₅	0,622	x ₂₃		0,917	
3-я группа	x ₁₃	0,743	x ₂₄	0,600	
	x ₁₇	0,718	x ₁	0,762	
	x ₂₀	0,678	x ₆	0,522	
	x ₃₁	0,694	x ₁₁	0,467	
4-я группа	x ₃₄	0,646	6-я группа	x ₁₈	0,727
	x ₂	0,669		x ₁₉	0,722
группа	x ₃	0,866	x ₂₃	0,510	
	x ₄	0,962	x ₃₀	0,770	
			x ₃₂	0,642	

При разбиении на 12 групп состав их и теснота связи оказались такими:

Фактор		r	Фактор		r
1-я группа	x ₁₃	0,882	5-я группа	x ₁₇	0,922
	x ₂₀	0,773		x ₃₄	0,922
	x ₃₁	0,798		6-я группа	x ₆
2-я группа	x ₇	0,774	x ₁₆		0,730
	x ₁₂	0,674	x ₃₃		0,772
	x ₂₉	0,636	7-я группа	x ₈	0,881
3-я группа	x ₁	0,845		x ₃₅	0,881
	x ₂	0,822		8-я группа	x ₁₁
	x ₃	0,951	x ₂₇		0,628
	x ₅	0,762	x ₂₈		0,683
	x ₂₁	0,938	9-я группа	x ₉	0,959
4-я группа	x ₄	0,967		x ₁₅	0,959
	x ₁₄	0,784			
	x ₂₇	0,919			
	x ₂₂	0,969			

10-я группа	x ₃₀	0,840	11-я группа	x ₁₈	0,762
	x ₃₂	0,692		x ₁₉	0,773
группа	x ₁₀	0,981	группа	x ₂₄	0,741
	x ₂₃	0,981		x ₂₆	0,741

Группы, сформировавшиеся при более дробном разбиении, служили основой для образования укрупненных групп. После этого конструировались различные варианты сокращенных наборов, для чего проводился целенаправленный отбор одного — двух признаков из каждой группы.

Представители группы выбирались на основе ряда соображений — как содержательных, так и формальных. С содержательных позиций предпочтительней были параметры, значения которых легко оценить для вновь проектируемых и конструируемых предприятий, например средний вес плавки — x₃, годовая выплавка стали — x₄ и др.

С формальной точки зрения заслуживали внимания:

а) расстояние признака до условного среднего признака группы — чем ближе признак к среднему, тем предпочтительнее его выбор;

б) величина парных коэффициентов корреляции между признаками и моделируемым показателем — чем больше теснота связей, тем предпочтительней выбор;

в) связь признаков с уже отобранными признаками — чем меньше связь, тем предпочтительней выбор;

г) состав группы при разных разбиениях — целесообразно включать в набор признаки, образующие при разных разбиениях единичные группы.

На основании приведенных соображений из 35 параметров, подвергнутых факторному анализу, для этапа непосредственного моделирования были оставлены 20:

- x₁ — емкость мульд, м³;
- x₂ — количество шихты, заваливаемое 1 машиной в год, т;
- x₃ — средний вес плавки, т;
- x₄ — годовая выплавка стали, тыс. т;
- x₅ — годовая выплавка стали с кислородом в ванну, % к общей выплавке с кислородом;
- x₆ — годовая выплавка низколегированной стали, %;
- x₇ — » » углеродистой » , %;
- x₈ — » » спокойной » , %;
- x₉ — доля чугуна в садке, %;
- x₁₀ — содержание в чугуне фосфора, %;
- x₁₁ — » » кремния, %;
- x₁₂ — » » куска в ломе, %;
- x₁₃ — количество стали, разлитой в изложницы на тележках, %;
- x₁₄ — количество стали, разлитой в изложницы сверху, %;

- x_{15} — расход окислителей, % к общей шихте;
- x_{16} — калорийность топлива, ккал;
- x_{17} — общие простои, пече-суток на 1 печь;
- x_{18} — средний срок действия цеха, лет;
- x_{19} — удельный расход топлива, кг условного топлива на 1 т;
- x_{20} — удельный расход кислорода, м³/т.

Глава 3

ТИПЫ СТАТИСТИЧЕСКИХ МОДЕЛЕЙ И МЕТОДЫ ИХ ПОСТРОЕНИЯ

§ 1. Выбор рационального типа модели

3.1. В практике экономико-статистического моделирования наиболее широко используются непрерывные линейные регрессионные модели. Они предполагают качественную однородность¹ рассматриваемой совокупности объектов, отсутствие разрывов и скачков в моделируемой зависимости и требуют нормального или близкого к нормальному многомерного распределения признаков, описывающих объекты. Принимается во внимание уменьшенный набор признаков, полученный на этапе предмодельного анализа исходной информации (гл. 2, § 4).

3.2. Пространственные совокупности промышленных объектов одной отрасли (подотрасли), как правило, неоднородны, они распадаются на ряд подсовокупностей со специфичным для каждой характером зависимости экономических показателей от входных переменных. Причиной этому служат различия в уровне техники, технологии и организации производства на отдельных объектах совокупности. Эти различия обусловлены неодинаковым сроком ввода объектов в эксплуатацию. Объекты, построенные в один или близкие календарные периоды, как правило, мало различаются по уровню техники и технологии производства и находятся на примерно одинаковых точках траектории своего развития. Объекты же разных «поколений» обычно существенно различаются между собой по своим

¹ Понятно, что речь идет не об абсолютной однородности, а об относительной, условной, при которой полученные статистические закономерности формирования экономических показателей справедливы для каждой единицы совокупности.

характеристикам и характеру влияния факторов производства на экономические показатели. Далее, объекты одного «поколения» могут, в свою очередь, распадаться на группы. Здесь в качестве типобразующих факторов выступают мощность объекта (как правило, крупные и мелкие предприятия оснащены разной техникой и имеют различный уровень организации производства и, как следствие, неодинаковые зависимости факторов и показателей), назначение и характер продукции, в добывающих отраслях — природные условия и т. д.

3.3. При изображении объектов системы в виде точек многомерного признакового пространства в большинстве случаев в расположении этих точек будут выделяться определенные области сгущений (типы объектов). В пределах области признакового пространства, занимаемого отдельным типом объектов, зависимость выходных переменных от входных носит непрерывный характер. При переходе от типа к типу изменяется и характер зависимостей, в отдельных случаях достаточно резко. Описание такой структурно-неоднородной совокупности объектов одним достаточно простым оператором дает значительные погрешности в индивидуальных оценках значений моделируемого показателя.

Если совокупность состоит из одного многочисленного класса и нескольких мелких, то на основной массе наблюдений погрешность сравнительно невелика. Когда же совокупность включает несколько «далеко» расположенных в признаковом пространстве классов с примерно равной численностью объектов и с различным характером зависимостей, тогда найденный по определенному критерию единый оператор будет лишен содержательного смысла, получим лишь какую-то формальную «среднюю» зависимость выхода от входа. Поэтому задача построения модели, адекватно отражающей существенные взаимосвязи выхода и входа объекта и приложимой к отдельному наблюдению, в условиях такой структурно-неоднородной совокупности объектов имеет первостепенное значение.

3.4. В более простых случаях структурной неоднородности адекватная ЭСМ строится с помощью подбора формы зависимости (используются нелинейные связи). Однако в общем случае это не решает проблемы по ряду причин. Во-первых, поиск необходимой формы связи для многофакторных моделей, по существу, ведется вслепую.

Во-вторых, значительное увеличение числа параметров приводит к снижению точности оценок и усложняет интерпретацию. В-третьих, в условиях неоднородности большинство статистических критериев не работает.

3.5. Наиболее радикальный способ учета структурной неоднородности промышленных объектов состоит в аппроксимации реальной зависимости некоторой кусочно-линейной поверхностью, обеспечивающей достаточный уровень точности оценки показателей для практических приложений. Здесь может иметь место синтез методов распознавания образов и корреляционно-регрессионного анализа.

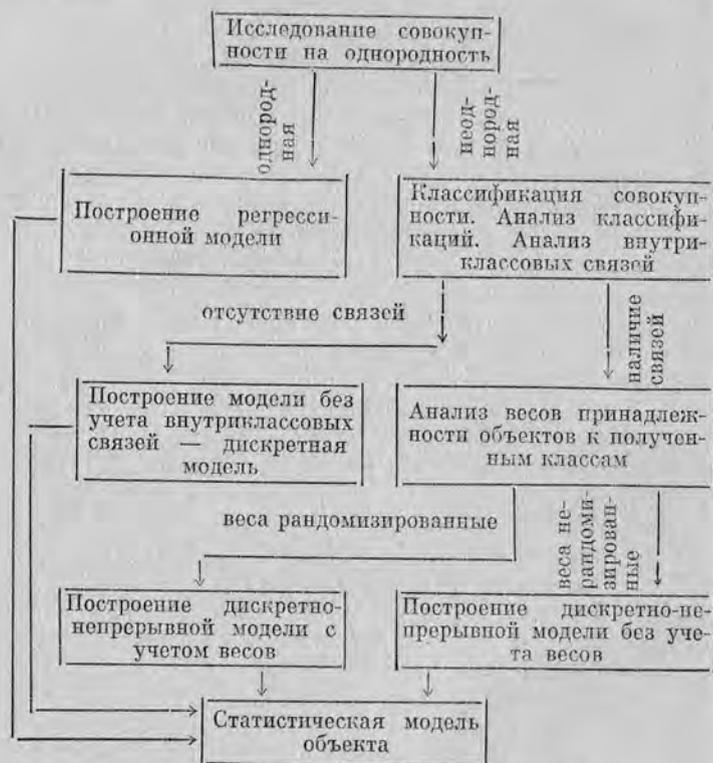


Схема 1. Построение статической модели объекта.

Построение статической модели объекта целесообразно проводить в такой последовательности (схема 1):

- 1) совокупность объектов исследуется на однородность;
- 2) если совокупность однородна, то строится непрерывная регрессионная модель;
- 3) неоднородная совокупность методами классификации разбивается на классы «близких» точек;
- 4) при отсутствии внутриклассовых связей (нет существенной вариации признаков и показателей по объектам, принадлежащим к одному классу) строится дискретная модель, описывающая связь классификаций — в пространствах факторов и показателей производства;
- 5) если внутриклассовые связи существенны, то для каждого из выделенных классов строится своя функция регрессии (с учетом весов принадлежности к классам или без) и совокупность этих регрессий есть дискретно-непрерывная модель (ДНМ).

§ 2. Исследование совокупности на однородность. Построение регрессионной модели

3.6. Исследование совокупности объектов на однородность практически тесно связано с возможностью построения адекватной линейной регрессионной модели, в то же время многомерное нормальное распределение всегда имеет своим условным математическим ожиданием линейное уравнение. Выделяются три типа методов выявления неоднородности совокупности.

Первый из них связан с проверкой гипотезы о нормальности многомерного распределения показателя и факторов производства. Известно, что если одномерное распределение не является нормальным, то не может быть нормальным и любое многомерное распределение, включающее первое в качестве одного из измерений. Поэтому, а также вследствие их сравнительной простоты указанные методы получили широкое распространение. К ним можно отнести построение и анализ полигонов частот или гистограмм для изучаемых признаков; проверку гипотез о согласии эмпирических однородных законов распределения с нормальным и близким к нему распределениями по различным критериям — Пирсона, Колмогорова, дифференциальному², приближенный метод оценки неизвестно-

² См., например, работы М. Кендалла, А. Стьюарта (1966), А. К. Митропольского (1971), М. Л. Лукацкой (Статистические методы в экономическом анализе производства. Новосибирск, 1968, с. 12—31).

го числа нормальных компонент, их параметров, а также величин, определяющих весá этих компонент в исследуемой смеси нормальных распределений (Bhattacharya, 1967).

Ко второму типу относятся методы, связанные с проверкой гипотезы о линейности регрессии: анализ парных корреляционных полей, анализ отклонений фактических значений от теоретических с помощью критериев знаков и серий и др. (Езекпел, Фокс, 1966; Уилкс, 1967; Ван дер Варден, 1960).

Третий тип составляют методы, основанные на разбиении совокупности на классы и сравнительном анализе основных статистических характеристик этих классов: сравнение соответствующих многомерных распределений (Заславский, 1969), проверка гипотезы о несущественности различия двух или нескольких регрессий (Бородюк, 1967; Chow, 1960; Gujarati, 1970), многомерная классификация с последующей оценкой существенности различия полученных в результате разбиения классов.

3.7. Если совокупность однородна, то строится непрерывная регрессионная модель: выбирается количество и набор существенных признаков, включаемых в модель; оцениваются неизвестные параметры и различные статистические характеристики — средние квадратические отклонения оцениваемых коэффициентов, остаточные дисперсии, отношения Фишера, матрица ковариаций отклонений и пр. Процесс завершается анализом и интерпретацией полученной модели. Построение регрессии для одного показателя обычным методом наименьших квадратов (м. н. к.) довольно широко освещено в литературе (Андерсон, 1963; Крамер, 1975; Кульбак, 1967; Рао, 1968; Юл, Кендалл, 1960; Езекпел, Фокс, 1966; Ван дер Варден, 1960; Кендалл, Стюарт, 1966; Митропольский, 1971; Дрейнер, Смит, 1973; и др.). В наших исследованиях использовались алгоритмы и программы, разработанные в ИЭ и ОПГ СО АН СССР (Статистические алгоритмы и программы, 1970; Алгоритмы статистической обработки информации, 1974).

3.8. Основные свойства оценок коэффициентов регрессии, полученных методом наименьших квадратов, следующие.

Пусть \vec{Y} вектор наблюдений (размерностью $n \times 1$) над выходной переменной y ;

\vec{X} — матрица независимых переменных x_{ij} , называемых входными, размерностью $n \times m$;

\vec{A} — вектор параметров, подлежащих оцениванию, размерностью $m \times 1$;

$\vec{\varepsilon}$ — вектор ошибок размерностью $n \times 1$.

Запишем в матричных обозначениях модель регрессии

$$\vec{Y} = \vec{X}\vec{A} + \vec{\varepsilon},$$

где $M(\vec{\varepsilon}) = 0$, $D(\vec{\varepsilon}) = I\sigma^2$, элементы вектора $\vec{\varepsilon}$ некоррелированы.

Оценка \vec{a} — вектора неизвестных параметров — может быть получена решением нормальных уравнений

$$\vec{a} = (\vec{X}'\vec{X})^{-1} \vec{X}'\vec{Y}'.$$

Это решение обладает рядом свойств.

I. Вектор \vec{a} есть оценка $\vec{\alpha}$, которая минимизирует сумму квадратов отклонений $\vec{\varepsilon} \cdot \vec{\varepsilon}$ безотносительно к тому, каков характер распределения ошибок.

II. Элементы вектора $\vec{\alpha}$ — линейные функции наблюдений y_1, y_2, \dots, y_n . Они представляют собой несмещенные оценки элементов $\vec{\alpha}$, обладающие минимальными дисперсиями безотносительно к характеру распределения ошибок.

III. Если ошибки независимы и $\varepsilon_i \approx N(0, \sigma^2)$, то \vec{a} есть оценка максимального правдоподобия величины $\vec{\alpha}$.

3.9. Необходимый этап построения непрерывных моделей — дальнейшее сжатие набора информационных признаков, полученного на этапе предмодельного анализа. Основные причины сокращения набора рассмотрены в гл. 2, § 4. Здесь укажем еще одну: доверительная область статистических характеристик при рассмотрении сравнительно небольшого числа наблюдений широка и точность такой модели мала. Сокращение набора признаков методом полного перебора невозможно с вычислительной точки зрения: даже при 10 факторах количество испытаний превышает 1000.

Известные в настоящее время алгоритмы отыскания группы существенных для данного показателя признаков неразрывно связаны с построением уравнений регрессии

показателя по различным группам признаков из полного набора и со сравнением их множественных коэффициентов корреляции или остаточных дисперсий. Наиболее точным представляется алгоритм случайного поиска с адаптацией. Однако он требует априорного знания о количестве существенных признаков и связан с большим объемом вычислений.

Другие алгоритмы основаны на последовательном исключении (или добавлении) признака, дающего наименьшее (наибольшее) изменение остаточной дисперсии по сравнению с уже взятой группой признаков. Исключение фактора проводится по величине коэффициента регрессии при нем.

Принципиальный недостаток этих алгоритмов заключается в том, что признак, исключенный на некотором шаге из группы значимых, не может быть возвращен в нее. Полученная таким образом на некотором шаге группа признаков не будет оптимальной для данного их числа. (То же самое относится и к добавлению признака.) Поэтому даже при использовании формальных критериев прекращения процесса выбора однозначно определить существенные признаки невозможно. Отметим также, что на каждом шаге алгоритмов отыскивается большое число уравнений регрессии, следовательно, объем вычислений очень велик.

Разработан сходящийся итерационный процесс построения уравнения регрессии по полному набору признаков методом последовательных приближений³. При использовании этого алгоритма для отбора группы существенных признаков некоторые упомянутые выше недостатки устраняются.

3.10. Выбор метода оценивания параметров зависит от вида модели. Если моделируется один показатель, то параметры модели оцениваются методом наименьших квадратов. При моделировании вектора показателей в большинстве случаев непосредственное применение м. н. к. для оценивания параметров каждого уравнения в системе приводит к несостоятельным и смещенным оценкам. Получить

³ Беккер А. В., Лукацкая М. Л. Алгоритм построения регрессионных моделей методом последовательных приближений. Выбор существенных факторов.— В кн.: Статистическое моделирование и прогнозирование технико-экономических показателей предприятий. Новосибирск, 1974, с. 147—154.

оптимальные оценки системы структурных уравнений позволяют специально разработанные способы, из которых наиболее прост двухшаговый м. н. к. (Лизер, 1971; Basman, 1957). Он предназначен для оценивания одного структурного уравнения в системе и учитывает только те ограничения, которые накладываются моделью на оцениваемое уравнение. Обобщением двухшагового м. н. к. на случай совместного оценивания системы структурных уравнений является трехшаговый м. н. к. (Madansky, 1964; Zelner, Theil, 1962; Basman, 1957). В отличие от двухшагового он относится к группе методов полной информации, так как в получаемых с его помощью оценках учитывается информация о ковариационной матрице структурных возмущений, а также ограничения, накладываемые моделью на остальные уравнения системы.

Оцениванию параметров системы структурных уравнений должно предшествовать исследование каждого ее уравнения на идентифицируемость. Понятие идентифицируемости уравнения в системе связано с наличием достаточного числа ограничений, налагаемых на него моделью. Если это условие не выполняется, то соответствующая оценка структурных уравнений невозможна.

Для идентифицируемости уравнения необходимо, чтобы соблюдалось условие порядка, которым обычно пользуются на практике. Если M — число эндогенных переменных в модели, K — число предопределенных переменных, H и I — соответственно число эндогенных и предопределенных переменных в исследуемом уравнении, то условие порядка записывается в виде

$$M - H + K - I \geq M - 1$$

или

$$K - I \geq H - 1.$$

3.11. Например, при моделировании производительности труда (y) на лесозаготовительных предприятиях из полного набора, содержащего 21 фактор, для конечного уравнения регрессии было отобрано 7: удельные веса в составе насаждений сосны (x_1), березы-осины (x_2), средний запас древесины на 1 га (x_3), среднее расстояние вывозки (x_4), объем вывозки (x_5), удельные веса вывозки древесины тракторами (x_6) и автомашинами (x_7). Уравнение регрессии имело вид

$$y = 420,5 + 1,922 x_1 - 0,581 x_2 + 0,045 x_3 - 3,014 x_4 + 0,214 x_5 + 0,15 x_6 + 1,495 x_7.$$

§ 3. Классификация совокупности

3.12. Если гипотеза об однородности совокупности отвергается, то необходима классификация, для чего выбираются набор признаков и способ классификации, формируется математическое описание классов, устанавливаются правила отнесения нового объекта к тому или иному классу и оценивается точность распознавания. В зависимости от целей моделирования и характера внутрикласовых связей классификация будет либо конечным этапом построения дискретной модели, либо промежуточным — для дискретно-непрерывной модели.

Всякая классификация преследует определенные цели — внешние, или, как их иногда называют, «суперцели», и внутренние. Под «суперцелью» обычно понимают тот формализованный содержательный критерий, которому должна отвечать классификация. Ведь нас удовлетворяет не всякая группировка, а лишь отражающая закономерности рассматриваемой задачи. К внутренним целям относятся «близость» объектов в группах, «удаленность» классов, «одинаковость» распределения объектов внутри класса и др.

3.13. Всякая классификация основана на интуитивном понятии «близости», или «сходства», объектов совокупности. Хорошим считаем такое разбиение выборки, при котором похожие объекты объединяются в один класс, а непохожие распределяются по разным классам. Если множество абстрактно, то введение понятия «близости» между его элементами соответствует заданию в нем некоторой топологии. Степень согласованности классификации с введенной топологией может показывать мера компактности, или критерий группировки. То есть для определения целевой функции надо задать топологию в пространстве описаний и критерий группировки.

Если топология определена на всем пространстве описаний и введен критерий группировки, то целевую функцию для произвольной выборки можно подсчитать по единой методике. При любых других способах задания целевой функции, зависящих явно от выборки, это практически невозможно. В свою очередь, топология пространства описаний может быть задана при помощи некоторой функции ρ , ставящей в соответствие каждой паре точек x, y из множества M вещественное число $\rho(x, y)$. В применении

к задачам классификации эту функцию называют мерой сходства, или близости, между объектами.

3.14. Далее нужно задать меру компактности группировки (более узкое понятие по сравнению с целевой функцией). При построении мер компактности можно исходить из нескольких концепций.

Рассеяние. Определим рассеяние в i -м классе — W_i и рассеяние между i -м и j -м классами — W_{ij} следующим образом:

$$W_i = \sum_{x, y \in A_i} \rho(x, y);$$

$$W_{ij} = \sum_{\substack{x \in A_i, y \in A_j, \\ i \neq j}} \rho(x, y).$$

Комбинируя их, можно получить меру компактности вида

$$Q_1(R) = \sum_{i=1}^S W_i - \sum_{\substack{i, j \\ i \neq j}} W_{ij}.$$

Если не учитывать межклассовое рассеяние, то придется вводить потери на число классов

$$Q_2(R) = \sum_{i=1}^S W_i + F(S).$$

Иногда целесообразно ввести дополнительные, связанные с разбросом классов по мощности, потери

$$Q_3(R) = Q_2(R) + \sum_{i=1}^S \left(|A_i| - \frac{n}{S} \right)^2.$$

Возможны и другие модификации меры компактности, продиктованные существом задачи.

Энтрoпия. В качестве меры рассеяния одного класса используется выборочная энтропия

$$H = \frac{1}{2} \sum_{x, y \in A_i} [\rho(x, y)/T] \ln [\rho(x, y)/T].$$

где T — фактор нормализации — равен

$$T = \frac{1}{2} \sum_{x, y \in A_i} \rho(x, y).$$

С помощью выборочной энтропии строится общая мера компактности либо как сумма энтропий всех классов (при

фиксированном числе классов), либо другим аналогичным образом.

Диаметр. В метрическом пространстве можно ввести понятие диаметра класса A_i

$$D_i = \sup_{x, y \in A_i} \rho(x, y)$$

и расстояния между i -м и j -м классами

$$D_{ij} = \inf_{x \in A_i, y \in A_j} \rho(x, y).$$

Комбинируя их, получаем меру компактности вида

$$Q_4(R) = \sum D_i - \sum_{i, j} D_{ij} \quad \text{и др.}$$

3.15. Из всего сказанного следует, что для автоматической классификации учитель в любом случае должен сообщить машине целевую функцию, так сказать, «обучить» ее цели классификации. Самопроизвольно машина сформулировать цель не может, поэтому термины «обучение» и «самообучение» относятся лишь к форме передачи информации о целевой функции учителем машине.

3.16. Результаты классификации полностью зависят от правильности формирования набора классификационных признаков. Состав его определяется внешней целью разбиения. Логика взаимоотношений «цель — критерий — набор» такова: целью диктуется директивное задание, набор выступает рабочим аппаратом, реализующим задание, а критерий выполняет контрольные функции. В соответствии с целью каждый классификационный признак должен отражать существенные, а не второстепенные отличия классов друг от друга. Можно рекомендовать такую последовательность формирования набора признаков:

а) составление максимального списка классификационных признаков;

б) первое сокращение списка на основе содержательных представлений о причинах, порождающих изучаемую неоднородность (знания теории явления, мнения экспертов);

в) второе сокращение списка в связи с проверкой обеспеченности информацией и ее качества;

г) третье сокращение списка с привлечением формальных методов статистического анализа вариационных рядов

классификационных признаков и их взаимокорреляционных функций, а также распознавательных методов отбора информативных признаков;

д) преобразование признаков с учетом требований, предъявляемых алгоритмом и программой — выбранным методом классификации.

В число признаков-кандидатов для формирования рабочего признакового пространства в первую очередь включаются качественные признаки. Отбор «претендентов» из количественных признаков требует предварительного исследования характера их распределения и связи с моделируемым показателем. Здесь следует обратить внимание на те количественные признаки, значения которых располагаются в определенных интервалах числовой оси, имеющих между собой разрывы (многовершинное распределение).

Полезен выбор классификационных признаков на основе анализа корреляционных полей и эмпирических линий регрессии моделируемых показателей от изучаемых факторов. В этом случае, исследуя градиенты, выявляют общую тенденцию зависимости на всем диапазоне изменения факторов и ее примерную форму. Затем анализируют расположение точек корреляционного поля внутри наиболее характерных интервалов изменения признаков и устанавливают частные тенденции для интервалов. Сравнение частных тенденций между собой и с общей тенденцией позволяет судить о наличии неоднородности зависимостей и силе этой неоднородности. Если частные тенденции составляют лишь отдельные участки общей, то предположение о неоднородности совокупности не оправдывается. В противном случае, когда между частными тенденциями какого-либо признака обнаруживаются явные различия, этот признак включается в классификационный набор.

3.17. Различают два способа классификации: последовательное логическое деление по каждому признаку в отдельности и формирование групп на основе набора признаков, ни один из которых не является необходимым или достаточным условием принадлежности к данной группе. Первый подход реализуется комбинационной группировкой или ее модификацией — методом последовательных разбиений, при втором используются различные алгоритмы классификации и распознавания образов.

Целесообразность применения того или иного способа

зависит от характера исходного материала. Если набор признаков-кандидатов включает в основном качественные признаки и близкие к ним, то целесообразен многошаговый процесс последовательных разбиений, когда на первом шаге совокупность разбивается по признаку с наибольшим весом неоднородности, на втором каждая из полученных частей разбивается по своему информативно-му признаку и т. д.⁴

Когда набор состоит из большого числа количественных признаков, предпочтительнее использовать алгоритмы классификации. Комбинационная группировка в этом случае ведет, во-первых, к разрушению естественно сложившихся групп объектов, во-вторых, к формированию множества малочисленных групп, не пригодных для статистического анализа.

В наших исследованиях использовались алгоритмы классификации, основанные на методе сферических решающих функций (Елкина, Загоруйко, 1966), потенциальных функций (Аркадьев, Браверман, 1971), непараметрических оценок плотности вероятности (Распознавание образов, 1972, с. 95—105) и соответствующие программы. Классификации совокупности объектов, полученные по различным алгоритмам, в принципе не совпадают. Более того, некоторые алгоритмы таковы, что классификация зависит от порядка в списке объектов. Заранее нельзя сказать, какой алгоритм лучше для того или иного практического исследования. Необходимо проанализировать полученные классификации и выбрать ту, которая лучше согласуется с профессиональными знаниями об изучаемой совокупности объектов.

Но одну рекомендацию можно сделать. Если классификация проводится для построения ЭСМ, т. е. нужно получить не просто компактные группы, а такие, в которых распределение признаков близко к многомерному нормальному, то лучше применить программу таксономии «Рельеф»⁵.

⁴ Журавель Н. М., Журавель Ф. А. Последовательная группировка на основе качественных признаков.— В кн.: Распознавание образов и регрессионный анализ в экономических исследованиях. Новосибирск, 1972, с. 45—49 (ротапринт).

⁵ Ягольшпер М. А. Программа таксономии «Рельеф».— В кн.: Алгоритмы статистической обработки информации. Новосибирск, 1974, с. 80—88 (ротапринт).

3.18. При любом методе разбиения важно определить рациональное число классов. С этой точки зрения алгоритмы классификации можно подразделить на две группы: в первой число классов заранее задается исследователем, во второй оно в явной форме не задается, а находится в процессе разбиения. Как правило, априорно определить точное число классов в конкретных исследованиях не представляется возможным. Поэтому применяют разбиение на различное число классов (при разных параметрах) и выбирают разбиение, в наибольшей мере удовлетворяющее целям исследования. Выбор рационального числа классов (классификации) основывается как на формальных, так и на содержательных принципах.

3.19. Большинство описываемых в литературе критериев качества классификации основано на определении различных характеристик, связанных с компактностью полученных групп (средняя близость классов, удаленность классов и т. п.).

В наших исследованиях применялись два критерия: информационный статистический критерий проверки гипотезы об однородности средних значений и корреляционных матриц (Верховская, 1971) и эвристический критерий выбора варианта классификации по матрице вероятностей принадлежности объектов к классам.

Первый критерий основан на проверке статистической гипотезы об однородности средних значений и корреляционных матриц нескольких независимых нормальных многомерных выборок. Он использован в исследовании классификации мартиновских цехов, работающих на твердом чугуе, полученной с помощью алгоритма «Рельеф» (табл. 3.1). Результаты применения информационного критерия к полученным классификациям следующие:

<i>h</i>	Табличный критерий	Информационный критерий	
	χ^2 -распределения	с равными ковариационными матрицами	с различными матрицами
0,3	106,395	90,933	75,714
0,4	74,468	152,452	45,226
0,5	23,685	145,637	101,518
0,6	23,685	147,481	99,892

При параметрах аппроксимации функции плотности вероятности, равных 0,5 и 0,6, значения обеих информа-

Таблица 3.1

Описание классификаций мартеновских цехов, полученных с помощью алгоритма «Рельеф»

Значение h	Число классов	Номер класса	Число объектов	Размах вариации (max и min) признаков				удельный расход топлива, кг усл. топ./т			
				годовой выплавка низколегированной стали, %	годовой выплавка углеродистой стали, %	содержание в чугуне кремния, %					
0,6	4	1	30	127	308	254	1020	33	40	8,67	10,97
		2	91	31	163	75	430	26	43	6,27	10,97
		3	2	81	83	65	66	33	36	2,8	3,00
0,5	5	1	31	127	308	254	1020	33	40	8,67	10,97
		2	88	31	140	75	430	26	43	7,94	10,97
		3	2	63	68	170	181	36	36	6,24	7,07
		4	2	81	83	65	66	33	36	2,8	3,0
0,4	8	1	4	228	237	884	915	38	39	10,42	10,97
		2	8	238	308	416	554	35	37	9,78	10,97
		3	4	149	153	985	1020	33	35	9,55	9,9
		4	15	127	194	296	521	33	40	8,74	10,38
		5	88	31	140	170	181	36	36	6,24	7,07
		6	2	63	68	75	430	26	43	8,2	10,97
		7	2	81	83	65	66	33	36	2,8	3,0
0,3	10	1	4	228	237	884	915	38	39	10,42	10,97
		2	8	238	308	416	554	35	37	9,78	10,97
		3	4	127	144	673	752	38	40	9,89	10,724
		4	4	149	153	985	1020	33	35	9,55	9,947
		5	10	172	194	355	521	33	40	8,74	10,38
		6	65	36	163	75	430	31	43	7,94	10,38
		7	24	31	87	87	369	26	32	9,56	10,97
		8	2	63	68	170	181	36	36	6,24	7,07
		9	2	81	83	65	66	33	36	2,8	3,0

ционных статистик, вычисленных по данному критерию, больше табличного нецентрального χ^2 -распределения (при уровне значимости $\alpha=0,05$). Это позволяет утверждать, что при $h=0,5$ и $h=0,6$ гипотезу об однородности средних значений следует отвергнуть (классы существенно различаются между собой), а при $h=0,3$ и $h=0,4$ принять (классы неразличимы). Так как в классификациях объектов при $h=0,5$ и $h=0,6$ число объектов в одних и тех же классах почти совпадает, то можно считать классификацию предприятий по данным четырем признакам при $h=0,5$ наиболее приемлемой.

Следует отметить, что полученный результат полностью совпал с экспертным выбором наиболее приемлемого разбиения изучаемой совокупности предприятий.

3.20. Эвристический критерий качества разбиения использует матрицу вероятностей принадлежности объектов к классам. На основе ее исследования делаются важные выводы относительно компактности классов, степени изолированности и т. п. Достаточно корректно определить приемлемые разбиения позволяют сравнительно простые характеристики: коэффициент диагональности Q , равный отношению суммы диагональных элементов матрицы к сумме всех ее элементов, и средняя величина неопределенности H , подсчитанная как сумма неопределенностей всех строк данной матрицы.

При сравнении классификаций мартеновских цехов, полученных при различных h , значения указанных двух эвристических характеристик оказались следующими:

h	Q	H
0,4	0,908	3,156
0,5	0,824	2,523
0,6	0,991	1,880

Исследование качества классификации по характеристикам Q и H показало, что, во-первых, выбор приемлемой классификации по значению средней величины неопределенности должен быть пороговым, поскольку H есть монотонно убывающая функция в зависимости от числа классов, во-вторых, значение величины диагональности также зависит от числа классов, но в отличие от первой характеристики не является монотонной функцией. Минимум Q достигается при $h=0,5$. Именно при этом значении h классификация будет наиболее приемлемой.

3.21. Содержательный анализ комплекса классификаций, полученных одним методом по одной и той же программе, включает:

1) изучение структуры классификаций (численность классов, сходство их по объектам) и выявление единичных объектов;

2) исследование качественного состава классификаций (сравнение по классам среднего уровня, вариации классификационных признаков и других характеристик);

3) для динамической информации дополнительно — изучение закономерностей образования во времени новых

классов, перехода объектов из класса в класс, стабильности состава во времени.

Анализ и сравнение классификаций в трех указанных направлениях с помощью выбранного критерия позволяют определить их пригодность и найти лучший вариант группировки.

3.22. В исследованиях по мартеновскому производству экспертный анализ выявил нецелесообразность прямого применения распознавательных методов для классификации всей совокупности цехов. Действительно, априори на основе качественных различий цехов определены следующие структурные закономерности исследуемой совокупности. По типу процесса (скрап и скрап-рудный) она отчетливо разделяется на две подсовокупности (в пространствах производственных факторов и экономических показателей), каждая из которых имеет специфический характер связей между показателями и факторами. Внутри указанных подсовокупностей есть самостоятельные группы, которые можно рассматривать как исторически сложившиеся и технологически обособленные типы мартеновских цехов. Причем подсовокупность цехов скрап-процесса более однородна. Практически в ней определяющим типобразующим параметром выступает способ разлива стали, формирующий две группы цехов: с разливкой в стационарные каналы и в подвижные изложницы. Для цехов скрап-рудного процесса можно выдвинуть гипотезу о существовании пяти качественно отличных типов в зависимости от сочетания объемов производства и вида топлива.

Анализ комплекса классификаций всей совокупности мартеновских цехов, полученных различными распознавательными методами, показал, что результаты классификаций не соответствуют выдвинутой гипотезе о структурных закономерностях общей совокупности. В частности, в один класс объединились цехи с разным типом процесса, хотя в состав классификационных признаков включались количественные эквиваленты качественного признака «тип процесса»; по ряду классов оказался значительным разброс значений классификационных признаков, приводящий к пересечению классов. Кроме того, анализ динамического аспекта классификаций не выявил приемлемых тенденций в динамике состава классов.

Наилучшие с содержательных позиций результаты были получены при комбинации двух способов разбиения: предварительного разбиения методом логического деления всей совокупности мартеновских цехов по признаку «тип процесса» (две подсовокупности) и последующего разбиения полученных подсовокупностей методами многомерной классификации.

§ 4. Построение дискретной модели

3.23. В каждом из полученных классов исследуются вариация признаков и внутриклассовые связи. Если вариация незначительна, то значения признаков можно считать приближенно равными средним по классам и строить дискретную модель. В этом случае внутриклассовыми связями пренебрегают ввиду их малой значимости. Модели дискретного типа сравнительно часто используются в планово-экономической работе. Это обусловлено пороговым характером реакции экономических систем на те или иные воздействия. Важно не любое изменение экономических показателей, а лишь такое, которое существенно скажется на результатах работы зависящих от данного объекта элементов экономической системы.

Практически наиболее часто дискретные модели применяются в случаях, когда сама экономическая задача требует группировки объектов (например, при построении системы групповых нормативов экономического стимулирования предприятий), или для измерения уровня выходного показателя используются качественные признаки (марка, сорт продукции), или, наконец, когда число объектов в классе мало.

3.24. В общем случае дискретная модель экономических показателей представляет собой две классификации объектов (в пространствах факторов и показателей производства) и описание связи этих классификаций. Такая модель позволяет, зная класс объектов в одном признаковом пространстве, с той или иной степенью точности определить классовую принадлежность объекта в другом признаковом пространстве.

К дискретной модели предъявляется два основных требования: достаточно тесной связи классификаций и «хорошего» качества разбиений в каждом признаковом пространстве. Слабая связь между двумя классификациями

еще не означает, что исходное признаковое пространство непригодно для построения модели. Может оказаться, что при изменении способа и алгоритма классификации будет достигнута значимая связь. Дискретные модели могут строиться методами «свободной», «принудительной» классификации и их комбинацией (Розин, 1973).

3.25. При «свободной» классификации оба разбиения (в пространствах факторов и показателей производства) строятся независимо друг от друга, а затем оценивается взаимосвязь между ними. Связь классификаций при этом подходе основана только на соответствующем выборе признаковых пространств.

Для проверки гипотезы о несущественности связи между полученными классификациями может быть использован критерий χ^2 -Пирсона и нормализованные коэффициенты взаимной сопряженности Пирсона и Чупрова, рассчитываемые на основе χ^2 , а также мера близости, предложенная Б. Г. Миркиным (1969).

Критерием качества дискретной модели служит средний показатель силы связи между составляющими ее двумя классификациями — чем он больше, тем выше прогностическая сила модели. Величина этого показателя зависит от элементов каждой из двух классификаций — от удачности выбора признаковых пространств и оптимальности разбиения в них.

Метод «свободной» классификации учитывает реально сложившееся распределение объектов в признаковых пространствах, но получение достаточно тесно связанных классификаций по этому способу очень трудоемко.

3.26. «Принудительная» классификация также предполагает два разбиения, из которых одно независимо (первичное), а второе подстраивается под него, зависит от первого. Взаимосвязь классификаций предусматривается самим способом построения, число классов в обоих признаковых пространствах одинаково. Наиболее часто такой способ моделирования применяется при описании результатов функционирования объекта одним экономическим показателем. В этом случае достаточно провести классификацию в пространстве факторов производства и затем для каждого полученного класса (таксона) определить среднее значение экономического показателя и его доверительные интервалы.

В общем случае, когда оба признаковых пространства многомерные, построение модели этим методом предполагает последовательное решение двух задач теории распознавания образов: группировки в независимом признаковом пространстве, результатом которой будет построение активной (обучающей) классификации, и построения вторичной классификации в другом признаковом пространстве, т. е. обучения распознаванию первичной классификации.

В качестве независимой обычно выступает классификация по экономическим показателям. Для оценки качества аппроксимации может использоваться один из двух критериев. Первый — это коэффициент связи классификаций при раздельном распознавании объектов в обоих признаковых пространствах по сформулированным в ходе обучения решающим правилам. Вторым критерием выступает суммарное число ошибок распознавания класса экономических показателей на основе описания объектов в пространстве факторов производства.

К достоинствам рассмотренного способа построения дискретных моделей следует отнести заведомо тесную связь между двумя классификациями. Но здесь возможно нарушение естественно сложившейся классификации в зависимом признаковом пространстве. При принудительном переходе от одной классификации к другой группы (типы) могут быть «разрезаны», нарушены.

3.27. При комбинированном способе еще до выполнения операции свободного разбиения пространство факторов производства (размерность и метрика) преобразуется на основе изучения информативности факторов для разбиения по показателям. При построении модели экономических показателей естественно требование получить такую классификацию объектов в пространстве факторов производства, которая была бы близка к классификации по экономическим показателям. А это может быть достигнуто соответствующим преобразованием признакового пространства.

Комбинированная схема построения дискретной модели включает следующие этапы:

1) свободную классификацию (таксономию) в пространстве экономических показателей (\vec{Y}) и построение на этой основе обучающей выборки (при этом доверительные ин-

тервалы групповых оценок показателей не должны перекрываться);

2) обучение распознаванию этих классов в пространстве факторов производства (\vec{X});

3) преобразование пространства факторов производства по результатам обучения (переход от \vec{X} к \vec{X}'). Преобразование заключается в минимизации описания и придании (при необходимости) «весов» отдельным признакам;

4) свободную классификацию (таксономию) в преобразованном пространстве факторов производства (\vec{X}');

5) описание и измерение связи классификаций в пространствах (\vec{Y}') и (\vec{X}').

Графически построение модели комбинированным способом представлено на схеме 2.

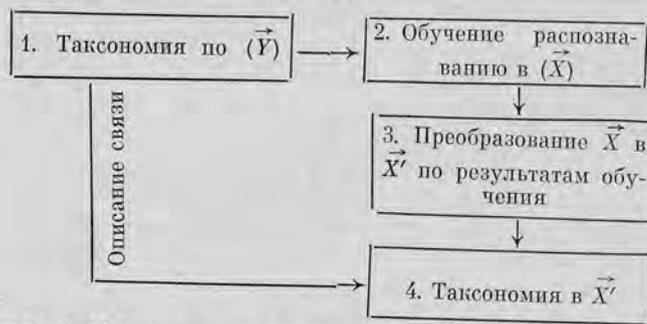


Схема 2. Построение дискретной модели (комбинированный способ).

Описанная многоэтапная схема гарантирует сравнительно быстрое и целенаправленное получение двух классификаций с достаточно тесной связью между ними (во всяком случае с близкой к максимально достижимой для данных признаков пространств). При этом сохраняется возможность выделения типов в каждом пространстве и проведения четких и простых границ между ними.

3. 28. Если для построения модели используется информация о положении объекта в двух признаковых пространствах, то при прогнозировании необходимо, зная положение объекта в одном признаковом пространстве, оценить его принадлежность к тому или иному классу в другом пространстве.

Основой для прогнозирования служит таблица взаимной сопряженности двух классификаций, составляющих дискретную модель. Для удобства использования при прогнозе в ряде случаев целесообразно от таблицы взаимной сопряженности, где

Таблица 3.2

Матрица частот

Классе	B_1	B_2	B_j	B_m	
A_1	P_{11}	P_{12}	P_{1j}	P_{1m}	$P(A_1)$
A_2	P_{21}	P_{22}	P_{2j}	P_{2m}	$P(A_2)$
A_i	P_{i1}	P_{i2}	P_{ij}	P_{im}	$P(A_i)$
A_n	P_{n1}	P_{n2}	P_{nj}	P_{nm}	$P(A_n)$
	$P(B_1)$	$P(B_2)$	$P(B)$	$P(B_m)$	

численности по классам представлены в абсолютном выражении $\|a_{ij}\|$, перейти к матрице частот $\|P_{ij}\| = \|a_{ij}/N\|$. Для этого все строки таблицы взаимной сопряженности делятся на общее число наблюдений (объектов) N . Форма матрицы частот приведена в табл. 3.2.

Качество прогнозирования зависит от вида матрицы. В идеальном случае число классов в обеих классификациях одинаково и каждому классу одной классификации соответствует только один класс в другой, т. е. имеется детерминированная связь; единицами в матрице заполнено лишь по одной клетке в каждой строке. В противоположном крайнем случае распределение частот в каждой строке равномерно, заполнены все клетки матрицы. Но тогда прогнозирование невозможно. В реальных ситуациях обычно встречаются матрицы промежуточного вида. В них каждому классу одной классификации соответствует несколько классов другой, однако построчное распределение вероятностей имеет вершину, и не все клетки матрицы заполнены.

3.29. Прогнозирование по матрице P_{ij} принадлежности объекта к классу B исходя из его принадлежности к классу A может выполняться в условиях, когда положение объекта в классификации A задано однозначно, т. е. известен класс A_i , в котором находится исследуемый объект (случай 1). Прогноз идет по схеме $A_i \rightarrow B_j$. Задача сводится к определению класса B_j , к которому принадлежит объект, находящийся в классе A_i . Для ее решения может быть использован критерий максимума условной вероятности.

Согласно теореме Байеса, можно записать соотношение

$$P(B_j/A_i) = P_{ij} \cdot P(B_j)/P(A_i),$$

где $P(A_i)$ и $P(B_j)$ — соответственно вероятности нахождения любого объекта в классах A_i и B_j .

Рассчитав значения $P(B_j/A_i)$ для строки A_i , находим класс B_j , соответствующий максимальному значению условной вероятности.

Возможна и другая ситуация: известна не классовая принадлежность нового объекта, а набор его признаков \vec{X}_k , на основе которых формировалась классификация A (случай 2). Прогноз выполняется по двухэтапной схеме — $\vec{X}_k \rightarrow A_i$ и $A_i \rightarrow B_j$. На первом этапе осуществляется распознавание класса объекта по его описанию, на втором — оценка классовой принадлежности в одном пространстве исходя из класса объекта в другом.

Если задача первого этапа решается в детерминистской постановке, то 2-й случай по технике прогнозирования сводится к 1-му. Особенность в прогнозировании возникает тогда, когда на первом этапе задача решается как вероятностная. В результате ее решения получаем распределение вероятностей принадлежности объекта \vec{X}_k к классу A_i — $P(A_i/\vec{X}_k)$.

В прогнозе в данном случае участвует не отдельная строка, а вся матрица P_{ij} , на основе которой рассчитывается математическое ожидание условной вероятности принадлежности объекта \vec{X}_k к классу B_j — $P(B_j/\vec{X}_k)$:

$$P(B_j/\vec{X}_k) = \sum_i P_{ij} P(B_j) \cdot P(A_i/\vec{X}_k) / P(A_i).$$

Определив значения математического ожидания условной вероятности для всех классов B_j ($j = \overline{1, m}$), можно соответственно максимальному из этих значений выбрать наиболее вероятный класс B_j для исследуемого объекта.

3.30. Для оценки прогностической силы модели обычно используется критерий суммарного числа ошибок распознавания при ретроспективных контрольных расчетах. При ограниченном размере исследуемой совокупности довольно часто в конкретных исследованиях используется метод скользящего контроля. Тогда построение модели и оценка ее прогностических возможностей выполняются многократно. Каждый раз для контроля из исследуемой

совокупности отбираются случайным образом один или несколько объектов, по оставшимся строится модель, прогностическая сила которой оценивается на отобранных объектах. Процедура повторяется достаточное число раз. Критерий точности модели определяется усреднением результатов по всем шагам итеративного процесса.

3.31. В исследовании мартеновского производства⁶ дискретная модель строилась по комплексу показателей («расходы по переделу», «итого задано в переработку сырья», «себестоимость мартеновской стали»). При построении модели классификация мартеновских цехов по «выходу» осуществлялась как по комплексу всех трех показателей, так и по различным сочетаниям их по два. Для классификации использовались алгоритм и программа, реализующие метод потенциальных функций.

Анализ полученных классов и процесса изменения их количественного и качественного состава при увеличении числа классов (2, 3, 4, 5) для различных наборов классификационных признаков показал, что классы наиболее компактны, однородны и стабильны при разбиении по двум показателям — расходам по переделу и себестоимости. Их и выбирают в качестве показателей «выхода».

Далее из множества факторов «входа» для дискретного моделирования формировались два набора. Минимальный включал 6 факторов (средний вес плавки, годовой объем выплавки стали, доля чугуна в садке, доля разливки стали в изложницы на тележках, калорийность топлива, удельный расход кислорода), а максимальный — 9 (к первому набору добавляются удельный расход топлива, процент выплавки низколегированной стали и содержание окислителей в шихте).

Сравнение соответствующих классов, полученных для этих двух наборов признаков, показало, что для всех классификаций область пересечения классов покрывает почти все пространство разбиений. Увеличение количества факторов в наборе практически не отразилось на характере разбиения совокупности, поэтому целесообразно в дальнейших расчетах использовать классификацию по 6-факторному набору. Наконец, для построения дискретной модели выявлялись связи между классификациями по «выхо-

⁶ Бондаренко Г. П., Журавель Н. М. Дискретное моделирование экономических показателей. — В кн.: Распознавание образов и регрессионный анализ в экономических исследованиях, с. 137—157.

ду» и «входу» и выбиралось некоторое наиболее «оптимальное» сочетание классификаций. Строились таблицы взаимной сопряженности двух классификаций, показавшие, что для большинства сочетаний классификаций имеет место статистически значимая связь между входными производственными факторами и изучаемыми комплексами выходных показателей. Для выбора рациональных параметров дискретной модели использовались два формальных критерия: максимальная величина коэффициента взаимной сопряженности между классификациями и максимальная ошибка прогноза, получаемая при проведении «экзамена» по вновь предъявляемым точкам.

Для всех классификаций были вычислены коэффициенты сопряженности Чупрова. Наиболее высокие значения этих коэффициентов получены для разбиения по показателям «расходы по переделу» и «себестоимость», причем максимум достигается в сочетании двухклассовых группировок как по «выходу», так и по «входу».

В процессе доводки модели до окончательного вида экзамен может преследовать две цели: проверку качества модели, степени ее адекватности моделируемому явлению и выбор оптимальной дискретной модели при наличии множества ее вариантов.

По каждой дискретной модели для экзаменуемых объектов подсчитывалась величина относительной ошибки отклонений фактических значений комплекса выходных показателей от тех, которые могут быть спрогнозированы. Среднее значение суммы этих ошибок по всем объектам принималось в качестве средней относительной ошибки прогноза по данной модели.

Для экзамена использовалась информация о работе 12 цехов. При наборе из шести входных факторов ошибка прогноза по двух- и трехклассовой моделям составила соответственно 5,5 и 5,2%. Дискретная трехклассовая модель приведена в табл. 3.3.

§ 5. Построение дискретно-непрерывной модели

3.32. Дискретно-непрерывная модель (ДНМ) — это комбинация моделей двух типов — дискретной и непрерывной. При ее использовании совокупность объектов моделируется на двух уровнях — классов и отдельных объектов. Модели верхнего уровня — дискретные, нижнего — непрерывные и представляют собой внутриклассовые

Таблица 3.3

Дискретная трехклассовая модель

Номер класса по «выходу»	Средние значения и средние квадратические отклонения входных факторов						Количество цехов, принадлежащих к классу j (i — номер класса по «входу», j — по «выходу»), и вероятность отнесения цеха к j -му классу, если известен номер класса по «входу»	
	средний вес плавления, т	годовая вылавка стали, тыс. т	доля чугуна в садке, %	доля разливки в изложницы на те-лежках, %	калорийность топлива, ккал	расход кислорода на 1 т мар-теновской стали, м ³		
1	202,004	1196,329	45,104	51,000	9,785	40,833	7 7 7 3	
2	439,401	1386,549	9,095	47,301	1,140	17,649	0,412 0,412 0,176	
3	445,859	578,965	41,953	54,929	9,279	6,046	9 15 4	
	90,011	698,896	12,966	45,588	1,909	13,830	0,321 0,536 0,143	
	234,243	1543,210	47,100	59,524	9,661	16,608	1 2 18	
	203,396	1460,111	12,452	47,857	1,751	21,804	0,018 0,095 0,857	
	Интервалы изменения значений выходного показателя						16,204—17,648	7,563—9,593
	Расходы по переделу, руб.						54,447—55,582	49,925—52,504
	Себестоимость 1 т мар-теновской стали, руб.						1 2 2	3
	Номер класса по «выходу»						1 2 2	3

уравнения регрессии. Исследуемая зависимость здесь представлена набором уравнений, каждое из которых действительно в определенной области признакового пространства. Дискретно-непрерывная модель имеет вид

$$y = \begin{cases} f_1(\vec{X}) & \text{при } \vec{X} \in A_1 \\ f_2(\vec{X}) & \text{при } \vec{X} \in A_2 \\ \dots & \dots \\ f_s(\vec{X}) & \text{при } \vec{X} \in A_s, \end{cases}$$

где A_1, A_2, \dots, A_s — полученные классы.

По сути, это развитие дискретной модели в случае, когда классы достаточно многочисленны и в них сохраняется существенная вариация показателя (y) в связи с вариацией факторов производства (\vec{X}). Дискретно-непрерывная модель целесообразна, если внутриклассовая регрессия существенно улучшает аппроксимацию моделируемого показателя по сравнению с внутриклассовой средней величиной, или там, где моделируется не один количественный показатель, а его функция от других переменных и времени.

Использование таких моделей эффективно для измерения качественных и количественных признаков в тех случаях, когда производственные факторы описываются двумя независимыми наборами признаков разной природы и каждый из этих наборов содержит существенную информацию о моделируемом процессе. Влияние качественных и близких к ним количественных признаков учитывается дискретной частью модели, остальных количественных — внутриклассовыми регрессиями. Качественные и близкие к ним количественные признаки оказывают весьма существенное влияние на моделируемый показатель либо при непосредственном воздействии их на уровень этого показателя, либо опосредованно, через воздействие на влияние других признаков. Дискретно-непрерывные модели позволяют учесть оба указанных аспекта влияния качественных признаков.

Процедура построения ДНМ включает следующие этапы:

формирование дискретной части модели⁷;

⁷ Этот этап описан в § 3 настоящей главы.

анализ компактности классов;

построение непрерывной части модели;

оценку аппроксимирующей и прогнозирующей способности всей модели.

3.33. Основу анализа компактности классов составляют веса принадлежности объектов к полученным в результате разбиения классам. Если совокупность состоит из S классов, это означает фактически, что каждому объекту Z_i соответствует S действительных, неотрицательных чисел $P_i^1, P_i^2, \dots, P_i^S$, где P_i^l — вероятность того, что объект Z_i относится к A_l -му классу. Естественно, что $\sum_l P_i^l = 1$

в случае, если i -й объект должен быть отнесен к одному из S классов, и $\sum_l P_i^l \leq 1$ в противном случае.

При «достаточно компактных» классах можно использовать нерандомизированные веса, а именно $P_i^l = 1$ для объекта Z_i , попавшего в класс с номером l , и $P_i^l = 0$ — для всех других классов. Когда гипотеза компактности нарушена (случай «близких» классов), возникает неопределенность при классификации граничных точек. Для учета такой неопределенности можно использовать рандомизированные веса (вероятности отнесения каждой точки к определенному классу).

Чтобы получить вероятность отнесения каждого объекта к каждому классу, по одному из алгоритмов строится оценка плотности вероятности $P_i(Z)$:

$$P_i(Z) = c_l \cdot \sum_{j=1}^{n_l} \exp \left\{ -\frac{1}{2} (Z - Z_j) H \sum_l^{-1} H' (Z - Z_j)' \right\},$$

где $H = \begin{pmatrix} h_1 & & 0 \\ & \ddots & \\ 0 & & h_{m+n} \end{pmatrix}$ — диагональная матрица весов признаков;

$$c_l = 1/n_l \cdot (2\pi)^{\frac{m+n}{2}} \cdot \prod_{i=1}^{m+n} h_i(n_l) \cdot \sqrt{|\sum_l|};$$

\sum_l — ковариационная матрица входных и выходных переменных класса A_l ; n_l — число объектов, попавших в класс.

Оценка априорных вероятностей каждого класса исчисляется как $P(A_l) = n_l/N$. Используя формулу Байеса, получим выражение для P_i^l — вероятности того, что объект

Z_i относится к l -му классу:

$$P_i^l = P(A_l) P_l(Z_i) \left| \sum_{j=1}^S P(A_j) P_j(Z_i) \right.$$

Заметим, что вероятности P_i^l можно находить, используя меру близости объектов и детерминированную классификацию. Пусть, например, для i -го объекта известен вектор $\rho_i = (\rho_{i1}, \dots, \rho_{iN})$ где ρ_{ij} — мера близости между точками Z_i и Z_j . Пронормируем его, т. е. получим вектор $\tilde{\rho}_i = (\tilde{\rho}_{i1}, \dots, \tilde{\rho}_{iN})$:

$$\tilde{\rho}_{ij} = \rho_{ij} / \sum_{j=1}^N \rho_{ij}, \quad i, j = \overline{1, N}.$$

В качестве вероятности P_i^l возьмем число

$$P_i^l = \sum_{z_j \in A_l} \tilde{\rho}_{ij}.$$

3.34. Построение непрерывной части модели включает все стадии, обычные для многофакторного регрессионного анализа. Специфика задачи проявляется лишь в особой важности стадии отбора существенных переменных, что связано с уменьшением численности выборки при разбиении совокупности на классы. Общие подходы к уменьшению размерности признакового пространства были рассмотрены в § 2 данной главы. При построении внутриклассовых регрессий целесообразно проанализировать уровень внутриклассовой вариации переменных. Те из них, которые имеют малую относительную вариацию, можно из рассмотрения исключить.

3.35. В случае хорошей делимости классов (выполняется гипотеза компактности) и наличия достаточно представительного набора объектов внутри выделенных подсовкупностей непрерывные модели в классах строятся без учета вероятностей принадлежности объектов к полученным классам (случай нерандомизированных весов) — обычными методами регрессионного анализа.

Если классы плохо делимы и нужно учесть неопределенность, возникающую вблизи соседних классов для оценки параметров внутриклассовых моделей, то критерием выбора этих параметров служит минимизация следу-

ющего функционала (остаточной дисперсии):

$$L_l = \sum_{j=1}^N \|y_j - y_l(x_j)\|^2 \cdot P_j^l.$$

Отсюда следует, что параметры каждой частной модели $y_l(x)$ в случае учета вероятностей отнесения объектов к полученным классам находятся методом наименьших квадратов со взвешенными наблюдениями.

3.36. Чтобы для некоторой точки x в пространстве факторов производства найти значение показателя y , необходимо знать вероятность отнесения данного объекта с параметрами x к одному из S известных классов, в каждом из которых есть уже своя функция регрессии. Эти вероятности определяются согласно рекомендациям, изложенным в § 3 настоящей главы, только в пространстве параметров x . Тогда значение показателя y для данного x может быть вычислено по формуле

$$y(x) = \sum_{l=1}^S y_l(x) P(l|x).$$

В качестве оценки адекватности полученной кусочно-линейной функции реальной зависимости предлагаем взять функцию

$$\sigma^2 = \sum_{i=1}^S P_i \sigma_i^2,$$

где σ_i^2 — оценка, вычисленная для i -го класса объекта, т. е.

$$\sigma_i^2 = M \{ (y - y_i(x))' \sum_i^{-1} (y - y_i(x)) \}.$$

3.37. Построение дискретно-непрерывной модели проиллюстрируем на примере решения задачи нормирования расхода материалов для синхронных генераторов мощностью до 100 кВт, выпускаемых отечественными предприятиями. Исследованием было охвачено 55 типоразмеров машин серий ЕС, ЕСС, ЕССМ, ЕСС5, ВМЗ-4,5 и др.

В качестве функций рассматривались следующие экономические показатели:

удельный чистый вес черного проката (кг/кВт) — y_1 ;
 удельная норма расхода черного проката (кг/кВт) — y_2 ;
 удельный чистый вес электротехнической стали (кг/кВт) — y_3 ;
 удельная норма расхода электротехнической стали (кг/кВт) — y_4 ;

удельный чистый вес обмоточного провода (кг/кВт) — y_5 ;
удельная норма расхода обмоточного провода (кг/кВт) — y_6 ;
общий вес на 1 кВт мощности (кг/кВт) — y_7 ;
полная себестоимость (руб.) — y_8 .

На основе анкетного опроса специалистов отрасли был сформирован первичный набор признаков, предположительно оказывающих влияние на удельный расход материалов. В этот набор вошли признаки:

количественные (мощность — x_1 , КПД — x_2 , объем производства — x_3 , толщина изоляции провода — x_4);

количественные с небольшим числом значений (скорость вращения — 3 значения, толщина электротехнической стали — 3 значения, диапазон рабочих температур — 6 значений, точность и способ регулирования — 4 значения);

бивариантные качественные (форма исполнения, число полюсов, серийность);

многовариантные качественные (марка электротехнической стали — 5 градаций, класс изоляции — 3 градации, номер завода-изготовителя — 3 градации).

В целях единообразия количественные признаки с небольшим числом значений и многовариантные качественные были преобразованы в бивариантные качественные; градации, принимающие одно значение для всех или большинства объектов, исключались. В результате этих преобразований кроме восьми показателей (y_1 — y_8) и четырех количественных признаков (x_1 — x_4) для описания объектов использовались следующие бивариантные качественные признаки:

скорость вращения — z_1 (1000 и 1500 об/мин);
марка электротехнической стали — z_2 (Э11 и Э22);
серийность — z_3 (крупно- и мелкосерийная);
диапазон рабочих температур $\pm 35^\circ$ — z_4 ;
диапазон рабочих температур $\pm 50^\circ$ — z_5 ;
точность регулирования $\pm 2\%$ — z_6 ;
точность регулирования $\pm 5\%$ — z_7 ;
точность регулирования ручная — z_8 ;
соответствуют признаку: класс изоляции «А», «В», «Н» — z_9 — z_{11} ;

число полюсов машины (4- и 6-полюсные) — z_{12} ;

номер завода-изготовителя — z_{13} .

Одним из алгоритмов⁸ приведенный набор признаков был сокращен. Структура матрицы связей исследовалась отдельно для количественных и качественных признаков; для первых — на основе матрицы парных коэффициентов корреляции, для вторых — матрицы мер близости (Миркин, 1969). Анализ таких преобразованных матриц позволял сделать следующие выводы.

1. Практически полностью дублируют друг друга пары показателей y_1 — y_2 , y_3 — y_4 , y_5 — y_6 ; следовательно, из шести перечисленных показателей нужно моделировать только три.

2. Относительно независимы признаки x_3 и x_4 .

3. Наиболее тесно связаны с моделируемыми показателями по удельному расходу материалов признаки x_1 и x_2 ; z_1 и z_{12} ; z_{11} и z_2 ; z_3 , z_{10} и z_{13} ; z_9 и z_8 ; z_7 , z_5 и z_6 .

В рассмотренном примере естественно было остановиться на дискретно-непрерывной модели, поскольку исходная информация была представлена признаками двух типов — качественными и количественными. Влияние качественных признаков и количественных с небольшим числом значений учитывается системой дискретных моделей, влияние же остальных количественных — частными регрессионными моделями. Дискретные части модели строились методом последовательных разбиений. Признаки для каждого последующего шага выбирались на основе расчета весов однородности⁹.

Для признака z_4 нет дублирующих, и он имеет один из малых весов однородности. Поэтому на первом шаге совокупность была разбита по этому признаку. В результате получилось две группы: K — 25 машин с диапазоном рабочих температур $+35^\circ$ — 10° и M — остальные объекты с другим диапазоном температур.

Затем выполнялось разбиение группы K по признаку z_1 , имеющему наименьший вес однородности для этой группы. На втором шаге процесса получили подгруппы K_a — 6 машин и K_b — 9 машин. Из группы M разбиением по признаку z_{13} (представляет группу [z_3 , z_{10} , z_{13}]) также получились две подгруппы — M_a ($N=17$) и M_b ($N=13$).

Учитывая значения других качественных признаков, которые не участвовали в разбиении, и принимая во внимание объем совокупности, остановились на следующей группировке: K_a , K_b , M_a . Остальные объекты были отнесены к каждой из этих выделенных трех групп с разными вероятностями (вероятность принадлежности (P) объектов к выделенным группам подсчитывалась исходя из коэффициентов близости этих объектов к центрам групп).

После формирования классов в каждом из них исследовались парные связи количественных признаков, включая показатели. Причем каждое наблюдение учитывалось со своим весом (вероятностью). Анализ полученных матриц показывает, что для моделирования y_2 , y_4 , y_6 , y_7 в классах I и III достаточно взять три признака: x_1 , x_3 , x_4 (соответственно мощность, объем продукции, толщина изоляции провода статора в % от диаметра), а в классе II — все четыре (добавляется признак КПД).

При построении модели себестоимости (y_8) в качестве дополнительных признаков рассматривались и показатели y_2 , y_4 , y_6 , y_7 .

Модель для класса I (18 наблюдений с вероятностью 1,0 и 15 с вероятностью меньше 1,0) приводится ниже (в скобках указаны средние квадратические отклонения соответствующих регрессионных коэффициентов:

⁸ Беккер А. В., Лукацкая М. Л. Об анализе структуры матрицы коэффициентов связи. — В кн.: Вопросы экономико-статистического моделирования и прогнозирования в промышленности. Новосибирск, 1970, с. 273—284 (ротапринт).

⁹ Беккер А. В. Построение весовых коэффициентов информативности признаков. — В кн.: Вопросы экономико-статистического моделирования и прогнозирования в промышленности. Новосибирск, 1970, с. 260—272 (ротапринт).

Функция	a_0	a_1	a_2	a_4	R^2
y_2	20,32	-0,151 (0,035)	-0,0002 (0,0003)	0,111 (0,213)	0,432
y_4	9,31	0,082 (0,012)	0,00006 (0,00009)	0,210 (0,072)	0,753
y_6	1,81	-0,015 (0,001)	-0,00001 (0,00001)	0,004 (0,009)	0,825
y_7	24,33	-0,194 (0,026)	-0,0002 (0,0002)	-0,176 (0,159)	0,659
y_8	$= 1352,68 + 114,507x_1 - 163,976x_2 - 738,78x_4 - 12,470y_2 - 5,959y_4 -$ $- 0,014y_6 + 16,195y_7;$ $R^2 = 0,679.$				

Для оценки норм расхода проката черных металлов, электротехнической стали и обмоточных проводов по ДНМ для некоторых новых типов машин необходимо определить вероятности принадлежности их к нашим классам. Три типа генераторов (ЕС-82-4С, ЕСС5-81-6М и ЕСС-94) с учетом мер близости к центрам классов были отнесены с вероятностями, равными 1, соответственно к I, II и III классам. Результаты расчетов (табл. 3.4) показывают, что точность расчета по методу ДНМ достаточна для определения укрупненных норм расхода материальных ресурсов при ориентировочном определении себестоимости. Но это не исключает разработку подетальных норм расхода материалов на стадии непосредственного производства генераторов.

Таблица 3.4

Оценка норм расхода проката черных металлов электротехнической стали и обмоточных проводов

Класс	Тип генератора	y_2		y_4		y_6	
		по ДНМ	по фактическим данным	по ДНМ	по фактическим данным	по ДНМ	по фактическим заводским нормам
I	ЕС-82-4С	15,531	15,382	11,307	10,7	1,349	1,223
II	ЕСС5-81-6М	17,214	17,2	11,131	10,3	1,769	1,602
III	ЕСС-91-4	17,280	16,9	10,274	9,2	1,181	0,989

Глава 4

НЕКОТОРЫЕ ВОПРОСЫ ПОСТРОЕНИЯ ДИНАМИЧЕСКИХ ЭСМ

4.1. В зависимости от характера исходной информации и предположений относительно того, что именно в данном конкретном случае считается устойчивым, стабиль-

Вид исходной информации

Номер года	Номер объекта	Значение показателей					
		«ВЫХОДНЫХ»			«ВХОДНЫХ»		
		$y^{(1)}$	$y^{(k)}$	$y^{(m)}$	$x^{(1)}$	$x^{(j)}$	$x^{(p)}$
.
.
.
1		$y_{1t}^{(1)}$	$y_{1t}^{(k)}$	$y_{1t}^{(m)}$	$x_{1t}^{(1)}$	$x_{1t}^{(j)}$	$x_{1t}^{(p)}$
.
.
t	i	$y_{it}^{(1)}$	$y_{it}^{(k)}$	$y_{it}^{(m)}$	$x_{it}^{(1)}$	$x_{it}^{(j)}$	$x_{it}^{(p)}$
.
.
n		$y_{nt}^{(1)}$	$y_{nt}^{(k)}$	$y_{nt}^{(m)}$	$x_{nt}^{(1)}$	$x_{nt}^{(j)}$	$x_{nt}^{(p)}$
.
.

Пространственное сечение в t-й момент времени, $t=1,2,\dots,T$

ным во времени (например, тенденция изменения показателя, либо взаимосвязь его с другими величинами, либо только набор взаимосвязанных величин и пр.), для конструирования динамических ЭСМ могут быть использованы различные математико-статистические методы.

4.2. Исходной информацией для построения динамической ЭСМ совокупности объектов выступают обычно сведения о величине «входных» и «выходных» характеристик каждого из объектов изучаемой совокупности за каждый год исследуемого периода. Эти данные удобно представить в виде таблицы с тремя входами (табл. 4.1).

4.3. Статистическое исследование пространственного сечения, т. е. сравнительный анализ деятельности разных предприятий в определенный период времени, позволяет построить частную статическую модель формирования экономических показателей производства (методы и примеры построения таких ЭСМ приведены в гл. 3).

Изучение временного сечения (реализации), т. е. анализ деятельности определенного предприятия в течение всего исследуемого периода, служит основой для построения частной динамической модели одного конкретного объекта.

Анализ всего информационного массива (пространственных и временных сечений) проводится обычно для построения общей динамической модели, отражающей и статические и динамические закономерности формирования экономических показателей. Однако при решении этой общей задачи необходимо построение и анализ частных (статических и динамических) моделей.

§ 1. Динамические модели отдельных объектов (частные динамические ЭСМ)

4.4. Динамические модели отдельного объекта в зависимости от цели исследования и характера исходной информации подразделяются на три вида:

- 1) построенные на основе анализа временных рядов отдельных показателей;
- 2) построенные на основе анализа многомерных временных рядов в случае выраженного тренда;
- 3) адаптационные, построенные на основе многомерных временных рядов (случай стационарного режима с дрейфом характеристик).

4.5. Для построения моделей первого вида применяются методы анализа одномерных временных рядов, основанные на анализе и экстраполяции предыстории экономических показателей без выявления причин их изменения во времени. В этом случае принимается предположение об устойчивости закона изменения показателя во времени.

При использовании методов анализа многомерных динамических рядов акцент делается на экстраполяцию зависимостей между моделируемыми показателями и производственными факторами. При этом предполагается устойчивость либо самих зависимостей во времени, либо законов изменения параметров этих зависимостей и пр.

4.6. Приемы исследования временных рядов отдельного показателя давно разработаны и широко распространены. Имеется довольно богатый формальный аппарат; выявление тенденций методом скользящих средних, наимень-

ших квадратов, линейного и нелинейного программирования (при наличии ограничений), прогнозирование случайной компоненты по авторегрессионной модели, выделение сезонных и циклических колебаний, методы экспоненциального сглаживания и гармонических весов и др. (Мот, 1966; Тидтнер, 1965; Хеннан, 1966; Brown, 1963; Hellwig, 1967). Эти методы и практические аспекты их применения доступно изложены в ряде книг, посвящих прикладной и методический характер (Френкель, 1972; Экономико-статистические исследования, 1969; и др.).

Однако их прямое использование возможно лишь тогда, когда динамика показателя носит стабильный характер и можно предположить ее сохранение в будущем. Такие предпосылки, как правило, с определенными допущениями, соблюдаются для интегральных показателей больших систем в целом (например, для спроса на массовую традиционную продукцию отрасли), но не для экономических показателей отдельного предприятия. Применение рассматриваемого приема ограничено и размерами рядов — в экономических исследованиях обычно можно набрать максимум 10—20 годовых реализаций. Не играя существенной самостоятельной роли, методы анализа динамических рядов выступают обязательным составным элементом всех других методов описания динамики.

4.7. В качестве динамической модели отдельного показателя широко применима авторегрессионная модель. Если стационарный в широком смысле дискретный процесс $Y(t)$ для любого целого t можно представить в виде

$$Y(t) = \sum_{j=1}^p \beta_j Y(t-j) + \varepsilon(t),$$

где $\{\varepsilon(t)\}$ — последовательность независимых случайных величин, то $Y(t)$ называется процессом авторегрессии p -го порядка (Хеннан, 1966).

Построение схемы авторегрессии для заданного динамического ряда представляет интерес не только в тех случаях, когда из предварительного анализа ясно, что процесс является схемой авторегрессии определенного порядка, но и при поиске простого преобразования, приводящего к процессу, близкому к последовательности независимых случайных величин. Оно особенно важно, если анализируются отклонения динамического ряда от найденной тенденции

или исследуются ряды отклонений фактических и расчетных данных.

Если в результате профессионального анализа удалось выяснить, что исследуемый процесс представляет собой определенную авторегрессионную схему (известны все процессы-аргументы), то остается оценить неизвестные параметры β_j и вычислить остаточную дисперсию по эмпирическому ряду динамики. Обычно же такой определенности нет, и гипотезы доопределяются при статистическом анализе исходного материала. В этом случае может быть принята следующая процедура¹ анализа случайного процесса $Y(t)$, заданного своим динамическим рядом $\{y_1, y_2, \dots, y_n\}$.

1. Находится несмещенная оценка $z = f(t)$ математического ожидания $m(t)$ процесса $Y(t)$, и изучается процесс $X(t) = Y(t) - f(t)$ отклонений от тенденции на основании ряда $\{x_t\} = \{y_t - f(t)\}$, $t = 1, 2, \dots, n$.

2. Сравнением оценок $\rho_j(s)$ — автокорреляционных функций при различных началах отсчетов j — проверяется выдвинутая на основании профессионального анализа гипотеза о стационарности процесса $X(t)$.

3. Определяются выборочные коэффициенты автокорреляции $r(s)$, $s = 1, 2, \dots$, анализируется эмпирическая нормированная корреляционная функция и на основании указанного и профессионального анализа выдвигаются определенные гипотезы относительно порядка авторегрессии и сечений $X(t - \alpha_i)$, включаемых в авторегрессионную схему. Еще раз подчеркнем, что некоторые из гипотез могут не отражать структуру процесса, но в то же время соответствующие уравнения будут служить хорошей аппроксимирующей функцией.

4. Определяются параметры уравнений, соответствующих выдвинутым гипотезам; исследуются значения полученных параметров и остаточных дисперсий; выдвигается определенная гипотеза H_0 о виде авторегрессии.

5. По эмпирическим данным $\varepsilon_t = x_t - \sum_{j=1}^p \beta_j x_{t-j}$, $t = p + 1, p + 2, \dots, n$ исследуются отклонения $X(t)$ от про-

¹ Эта процедура и ряд других вопросов авторегрессионного анализа подробно изложены в книге «Экономико-статистические исследования промышленного производства» (1969).

цесса, соответствующего гипотезе H_0 . Если эти отклонения можно считать не автокоррелированными, то гипотеза принимается; в противном случае следует вернуться к п. 4.

6. Находится, если это возможно, функция распределения $F(x)$ случайной величины ε .

7. Исходя из равенства $X(t) = Y(t) - f(t)$ и уравнения, соответствующего принятой гипотезе H_0 , для $Y(t)$ записывается равенство типа

$$Y(t) = \sum_{j=1}^p b_j Y(t-j) + \varphi(t) + \varepsilon$$

(ε — случайная величина с найденной функцией распределения $F(x)$), выступающее моделью или аппроксимирующей функцией исходного процесса $Y(t)$.

По такой схеме отдельно отыскиваются параметры тенденций (функции $\varphi(t)$) и авторегрессионной схемы (b_j). Можно оценивать параметры функции вида

$$Y(t) = \sum_{j=1}^p b_j Y(t-j) + \sum_{i=0}^k a_i t^i + \varepsilon$$

сразу, применяя метод наименьших квадратов (Мот, 1966). При этом необходимо либо заранее задать вид функции, либо поступиться четкостью интерпретации.

4.8. Приведем пример построения авторегрессионной модели стоимости промышленно-производственных фондов завода по данным об их величине за 7 лет — с июня 1-го года по март 7-го года (табл. 4.2). Данные за последние 12 месяцев использованы только для статистической проверки полученных результатов.

Проведем авторегрессионный анализ по предлагаемой схеме.

1. Среднее значение последовательности y_1, y_2, \dots, y_{58} , среднее квадратическое отклонение и коэффициент вариации. равные соответственно 5,106 (млн. руб.); 0,222 (млн. руб.) и 4,3 (%), мало отличаются от аналогичных характеристик для полного ряда, составляющих 5,06 (млн. руб.); 0,226 (млн. руб.) и 4,5 (%) соответственно. Характер исходных данных, динамика средних значений по годам и небольшой коэффициент вариации показывают, что относительно исследуемого ряда может быть выдвинута гипотеза об отсутствии тенденций. В то же время видно, что колебания вокруг среднего значения не случайны: вначале

Таблица 4.2

Динамика промышленно-производственных фондов

Месяц	1-й год	2-й год	3-й год	4-й год	5-й год	6-й год	7-й год
I	—	4,80	5,25	5,08	5,08	4,86	4,78*
II	—	4,81	5,17	5,07	5,08	4,78	4,83*
III	—	4,80	5,17	5,08	5,09	4,84	5,00*
IV	—	4,81	5,31	5,19	5,06	4,85*	—
V	—	4,80	5,34	5,21	5,09	4,85*	—
VI	4,73	5,06	5,46	5,28	5,05	4,94*	—
VII	4,72	5,15	5,51	5,35	5,13	4,88*	—
VIII	4,77	5,25	5,56	5,38	5,12	4,82*	—
IX	4,88	5,21	5,48	5,30	5,09	4,81*	—
X	4,92	5,25	5,43	5,30	5,02	4,77*	—
XI	4,84	5,32	5,40	5,22	5,00	4,72*	—
XII	4,76	5,28	5,16	5,11	4,90	4,72*	—
Среднее значение	4,80	5,06	5,35	5,21	5,06	4,82	—

* Данные, использованные для статистической проверки полученных результатов.

значения ряда систематически меньше среднего арифметического, затем превышают его и лишь в конце 1964 г. становятся меньшими. Поэтому перейдем к анализу преобразований авторегрессионного типа, предполагая, что оценочной математического ожидания будет функция $f(t) = 5,106$.

2. Проверка гипотезы о стационарности процесса сравнением значений $\rho_j(s)$, $j=1, 2, \dots$ (коэффициенты автокорреляции s -го порядка, полученные по временному ряду, начиная соответственно с j -го его члена) показала, что нет оснований ее отбрасывать.

3. Оценки коэффициентов автокорреляции $r(s)$ ($s \leq 7$) для ряда отклонений от среднего значения $x_t = y_t - 5,106$, $t=1, 2, \dots, 58$ были следующие:

s	1	2	3	4	5	6	7
$r(s)$	0,923	0,816	0,685	0,540	0,425	0,325	0,268

С увеличением расстояния между сечениями корреляционная функция убывает, достигая минимального значения ($-0,625$) при $s=18$, затем возрастает и для $s \geq 20$ остается незначимой. В данной ситуации нецелесообразно учитывать такое длительное запаздывание (18 месяцев),

но окончательное решение принимается после исследования следующих гипотез о характере авторегрессионных преобразований:

$$H_1: X(t) = \beta_{11}X(t-1) + \varepsilon_1(t),$$

$$H_2: X(t) = \beta_{21}X(t-1) + \beta_{22}X(t-2) + \varepsilon_2(t),$$

$$H_3: X(t) = \beta_{31}X(t-1) + \beta_{32}X(t-2) + \beta_{3,18}X(t-18) + \varepsilon_3(t),$$

$$H_4: X(t) = \beta_{41}X(t-1) + \beta_{4,18}X(t-18) + \varepsilon_4(t).$$

4. Оценки параметров β_{jk} и множественных коэффициентов корреляции R_j (табл. 4.3), найденные методом наименьших квадратов соответственно выдвинутым гипотезам H_j , $j=1, 2, 3, 4$, показывают, что гипотезы о частичных авторегрессионных схемах следует отбросить, так как ко-

Таблица 4.3

Оценки параметров авторегрессионных преобразований

Гипотеза	Оценка			
	β_{j1}	β_{j2}	β_{j18}	R_j
H_1	0,923	—	—	0,923
H_2	1,139	-0,233	—	0,928
H_3	1,033	-0,179	-0,093	0,926
H_4	0,867	—	-0,090	0,921

эффициенты $\beta_{3,18}$ и $\beta_{4,18}$ незначительны. Кроме того, R_3 и R_2 , R_4 и R_1 незначимо различаются между собой. Иначе говоря, учитывать запаздывание в полтора года не имеет смысла. Для дальнейшего исследования оставим гипотезу H_1 об авторегрессии первого порядка.

5. Гипотеза об авторегрессии первого порядка была проверена способом, предложенным Кенуейм (Тинтнер, 1965, с. 331), и принята. Выяснили, что процесс $X(t)$ — авторегрессия первого порядка

$$X(t) = 0,923X(t-1) + \varepsilon,$$

где ε — случайная величина с нулевым математическим ожиданием и средним квадратическим отклонением, равным 0,086 (млн. руб.).

6. Анализ отклонений показал, что случайную величину ε можно считать распределенной нормально с параметрами $(0; 0,086)$.

7. Переход к $Y(t)$ дает авторегрессионную модель стоимости промышленно-производственных фондов в виде $Y(t) = 0,395 + 0,923 Y(t-1) + \varepsilon$, $\varepsilon = N(0; 0,086)$. (4.1)

4.9. Исследуем приемлемость полученной авторегрессионной модели для прогноза. Результаты прогноза по ней (за последние 12 месяцев) сравним с прогнозом по параболической модели, полученной методом наименьших квадратов:

$$y_2(t) = 4,605 + 0,0453t - 0,000724t^2. \quad (4.2)$$

Отклонения от фактических значений y_t центров прогнозируемых интервалов, подсчитанных по параболе (4.2) и по авторегрессионной схеме (4.1), т. е. отклонения точечного прогноза от действительных значений показателя, приведены в табл. 4.4.

Прогноз по авторегрессионной схеме проводился двумя способами: прогноз на следующий месяц $\tilde{y}_1(t)$, когда в ка-

Таблица 4.4

Сопоставление прогнозов по параболической тенденции и по авторегрессионной схеме первого порядка

Месяц	t	y_t	$y_2(t)$	$\tilde{y}_1(t)$	$\tilde{y}_2(t)$	$y_2(t) - y_t$	$\tilde{y}_1(t) - y_t$	$\tilde{y}_2(t) - y_t$
IV	59	4,85	4,76	4,86	4,86	-0,09	0,01	0,01
V	60	4,85	4,72	4,87	4,88	-0,13	0,02	0,03
VI	61	4,94	4,68	4,87	4,90	-0,26	-0,07	-0,04
VII	62	4,88	4,63	4,95	4,92	-0,25	0,07	0,04
VIII	63	4,82	4,59	4,90	4,94	-0,23	0,08	0,12
IX	65	4,81	4,54	4,84	4,95	-0,27	0,03	0,14
X	65	4,77	4,49	4,83	4,96	-0,28	0,06	0,19
XI	66	4,72	4,44	4,80	4,98	-0,28	0,08	0,26
XII	67	4,72	4,39	4,75	4,99	-0,33	0,03	0,27
I	68	4,78	4,34	4,75	5,00	-0,44	-0,03	0,22
II	69	4,83	4,29	4,81	5,01	-0,54	-0,02	0,17
III	70	5,00	4,23	4,85	5,02	-0,77	-0,15	0,02

честве $Y(t-1)$ выступали эмпирические значения показателя y_{t-1} , и далекий $\tilde{y}_2(t)$, когда значения показателя за последние 12 месяцев не принимались в расчет, а значением $Y(t-1)$ был центр прогнозируемого интервала $\tilde{Y}_2(t-1)$, подсчитанный по авторегрессионной схеме².

Прогноз по параболе оказался очень плохим, причем не только на длительный период: уже третье значение отличается от подсчитанного по параболе на 0,26, что составляет 2,23 среднего квадратического отклонения ряда $y_t - y_2(t)$, $t=1, 2, \dots, 58$. Из двенадцати отклонений только три меньше, чем удвоенное среднее квадратическое отклонение.

Прогноз на следующий месяц по авторегрессионной схеме дает хорошее совпадение с действительными значениями стоимости производственных фондов. Из двенадцати отклонений (предпоследний столбец табл. 4.4) только последнее превышает среднюю квадратическую ошибку прогнозирующей функции, причем менее чем в два раза. Далекий прогноз по авторегрессионной схеме (последний столбец таблицы) хуже, чем краткосрочный, но значительно лучше прогноза по параболе: только три отклонения (из двенадцати) превышают число 0,20, тогда как для параболы таких отклонений десять.

Проведенный анализ показывает, что динамической моделью авторегрессионного типа стоимости промышленно-производственных фондов можно пользоваться для прогноза по крайней мере на ближайший год. По мере появления новых данных необходима корректировка динамической модели с учетом дополнительной информации.

4.10. Один из методов построения динамической модели объекта по многомерным временным рядам (случай явно выраженного тренда)³ состоит в конструировании зависимости временного ряда показателя от временных рядов факторов производства («выхода» от «входа») с учетом их сдвигов во времени (лагов), а также от фактора времени. Этот тип модели отличает учет причин изменения показателя и хорошая приложимость к конкретному объекту.

² В этом случае среднее квадратическое отклонение прогнозируемой величины возрастает по мере увеличения прогнозируемого периода.

³ Литературные материалы по методам построения моделей на основе анализа многомерных рядов более скудны по сравнению с освещением анализа одномерных рядов.

Особо следует выделить возможность выявления при анализе многомерных временных рядов величины запаздывания во времени действия отдельных производственных факторов, что играет большую роль при любых способах построения динамических моделей.

С помощью этих же моделей можно учесть динамику основных экономических показателей, вызываемую влиянием фактора времени (вернее, факторами, стоящими за ним). Поэтому динамические модели отдельного объекта (помимо их самостоятельного применения) могут оказаться полезными и как вспомогательный элемент построения общей динамической модели совокупности объектов. Общим недостатком временных рядов как основы построения ЭСМ обычно заключается в ограниченности их длины, обусловленной как методологическими причинами (несопоставимость), так и организационными факторами (изменение характера отчетности, системы управления и т. д.). Это значительно уменьшает достоверность выводов по моделям. Второе ограничение на использование моделей рассматриваемого типа — узость диапазона изменения факторов и показателей производства на отдельном объекте, что, в свою очередь, существенно ограничивает сферу их применения.

4.11. Фактор времени своеобразен, он аккумулирует (опосредует) в себе влияние целого комплекса реальных причин развития процесса (измеримых и неизмеримых). В этом плане можно провести аналогию между фактором времени и географическим положением предприятия, за которым (особенно в добывающих отраслях) также стоит целый комплекс реальных факторов производства. Нагрузка на фактор времени в моделях во многом зависит от их структуры: чем больше факторов развития объекта удается ввести в модель в явном виде, тем эта нагрузка меньше. Однако полностью устранить влияние этого фактора невозможно, так как в нем выражается действие ряда неизмеримых факторов.

Время в моделях экономических показателей предприятий может выступать в двух формах: в календарной, и тогда начальная точка отсчета одинакова для всех объектов совокупности; и в форме «возраста» предприятия, времени его функционирования — начальные точки отсчета для объектов с разными сроками ввода в эксплуатацию неодинаковы.

Между двумя формами измерения времени существует тесная взаимосвязь, однако их смысловые нагрузки различны. Календарное время, введенное в модель, попросту учитывает в себя все, что связано с тенденцией развития моделируемого объекта и не нашло явного отражения в независимых переменных модели. По «возрастному» фактору предприятия с одинаковым значением календарного времени разделяются на разные «поколения». Это позволяет в неявной форме учитывать многие качественные моменты, отличающие предприятия с разным сроком ввода в эксплуатацию. Время функционирования предприятия — характеристика комплексная. Непосредственно она отражает процесс «старения» объекта, косвенно — технико-организационный уровень предприятия. При моделировании можно попытаться разделить эти две сферы действия фактора «возраста», для чего ввести в модель показатели года ввода объекта в эксплуатацию и его относительного «возраста».

4.12. Этапы построения линейной динамической модели отдельного объекта на основе анализа многомерных рядов динамики $y_t, x_{1t}, \dots, x_{kt}, t=1, 2, \dots, T$ рассматривались в гл. 1 и 2. Здесь дополнительно укажем еще три этапа: определение временных лагов на основе технико-экономического анализа и анализа корреляционных и взаимных корреляционных функций;

выбор зависимости изучаемого показателя от факторов-аргументов с учетом временных лагов, времени и значений показателя в предшествующие периоды времени, т. е. выбор основного уравнения модели;

нахождение системы уравнений, дополняющей основное уравнение до динамической модели.

Качественный анализ причинно-следственных связей показателя y с факторами производства $x_i, i=1, 2, \dots, k$, количественный анализ корреляционных и взаимных корреляционных функций, проведенный на первых двух вышеуказанных этапах, дают возможность записать процесс изменения исследуемого показателя $Y(t)$ в виде

$$Y(t) = \sum_{j=1}^p \beta_{ij} Y(t-j) + \sum_{i=1}^k \sum_{j=0}^{p_i} \gamma_{ij} X_i(t-j) + f(t) + \varepsilon. \quad (4.3)$$

То есть кроме основных $(k+1)$ признаков $Y(t), X_i(t)$,

$i=1, 2, \dots, k$ дополнительно в систему включаются и рассматриваются как самостоятельные признаки с запаздыванием $Y(t-j)$ ($j=1, \dots, p$), $X_i(t-j)$ ($j=1, \dots, p_i, i=1, \dots, k$) и фактор времени t . К основному (по смысловой нагрузке) уравнению (4.3) для получения линейной динамической модели необходимо присоединить систему таких уравнений, которые давали бы возможность по начальным условиям найти величины показателей в последующие моменты времени — третий этап схемы. Это может быть система типа

$$X_\lambda(t) = \sum_{i=1}^k \sum_{j=1}^{p_i+1} \gamma_{ij}^{(\lambda)} X_i(t-j) + \sum_{m=1}^{p+1} \beta_m^{(\lambda)} Y(t-m) + f_\lambda(t) + \eta_\lambda, \quad \lambda = 1, 2, \dots, k. \quad (4.4)$$

Следующий этап — оценка параметров системы уравнений по исходной информации. Он аналогичен рассмотренному в гл. 1, 2 и проводится либо обычным методом наименьших квадратов, если остатки ε , h_λ , $\lambda=1, 2, \dots, k$ слабо коррелированы, либо с помощью двух- или трехстадийного м. н. к. оценки системы совместных уравнений (Лизер, 1974; Basman, 1957; Madansky, 1964; Zelner, Theil, 1962).

Значения $Y(t)$ в любой момент для системы (4.3), (4.4) вычисляются по следующей схеме.

1. По «начальным условиям», т. е. по значениям $Y(t)$, $X_\lambda(t)$, $\lambda=1, \dots, k$ при $t \leq l_0$, из (4.4) находим $X_\lambda(l_0+1)$, $\lambda=1, 2, \dots, k$.

2. По «начальным условиям» и $X_\lambda(l_0+1)$ из (4.3) определением $Y(l_0+1)$.

3. Значения $Y(t)$, $X_\lambda(t)$, $\lambda=1, 2, \dots, k$ при $t \leq l_0+1$ можно снова считать начальными, отыскать по пп. 1 и 2 $Y(l_0+2)$ и т. д.

4. 13. Приведем пример построения динамической модели формирования себестоимости цемента по многомерному динамическому ряду информации о работе одного завода в течение 67 месяцев.

В результате предварительного анализа исходной информации были выделены факторы, влияющие на себестоимость цемента: объем производства, производительность труда, средняя марка, стоимость производственных фондов, ввод добавок, процент освоения мощности, численность промышленно-производственного персонала, расход электроэнергии, часовая производительность, коэффициент использования печей, влажность и тонкость помола шлама.

ма. В качестве аргумента был взят также параметр t — номер месяца. Исследование взаимокорреляционных функций показало целесообразность дополнительного включения некоторых факторов со следующим запаздыванием (в мес.):

Фактор производства	Запаздывание
Объем производства	6
Средняя марка	3
Стоимость производственных фондов	6
Часовая производительность	6
Тонкость помола шлама	3

Анализ регрессионных уравнений с различным набором факторов производства позволил установить, что от наличия или отсутствия аргумента t существенно изменяется только коэффициент при стоимости производственных фондов, ряды отклонений от регрессии при этом почти не изменяются. Для иллюстрации ниже приведены коэффициенты нормированных регрессий одной такой пары уравнений:

x_2	x_4	x_5	x_9	x_{10}	x_{13}	t	Стоимость производственных фондов
-0,238	-0,497	-0,091	0,235	-0,230	0,370	0,282	-0,380
-0,182	-0,484	-0,110	0,213	-0,239	-0,325	—	-0,132

Очевидно, что в равной мере можно пользоваться обоими уравнениями, но второе имеет то преимущество, что не требует знать начало отсчета для аргумента t .

При построении основного уравнения модели себестоимости желательно по возможности уменьшить количество аргументов. С этой целью найдены и проанализированы параметры линейных приближений себестоимости различными группами факторов. В частности, для некоторых из них множественный коэффициент корреляции и коэффициенты нормированных уравнений регрессии себестоимости 1 т цемента были следующими:

R	β_2	β_3	β_4	β_5	β_9	β_{13}	t
0,728		-0,145	-0,476	-0,417			0,361 0,317
0,718		-0,142	-0,488				0,275—0,081
0,729—0,155			0,465—0,400		-0,018		0,343 0,300
0,722—0,192			0,494		-0,036	0,098	0,260—0,126
0,722			-0,460—0,413				0,374 0,191

Множественные коэффициенты корреляции практически равны, коэффициенты при x_4 , x_5 и x_{13} устойчивы и имеют значительную величину. Поэтому в качестве основного уравнения модели себестоимости взята зависимость ее от средней марки (x_4), стоимости промышленно-производственных фондов (x_5) и тонкости помола шлама (x_{13}) с соответствующими запаздываниями:

$$X_1(t) = 12,20 - 0,0064 X_4(t-3) - 0,022 X_5(t-6) + 0,45 X_{13}(t-3). \quad (4.5)$$

Системой, дополняющей основное уравнение до динамической модели, может быть

$$X_4(t) = 674,76 - 25,378 X_1(t) + 0,4510 X_4(t-3);$$

$$X_5(t) = 1,365 - 0,0822 X_1(t) + 0,8770 X_5(t-6); \quad (4.6)$$

$$X_{13}(t) = -2,78 + 0,2209 X_1(t) + 0,0025 X_4(t-3) + 0,0021 X_5(t-6) + 0,4826 X_{13}(t-3).$$

Здесь каждое уравнение получено методом наименьших квадратов.

4.14. Специфическая структура системы уравнений, образующих динамическую модель себестоимости цемента (4.5)–(4.6), дает возможность прогнозировать на каждый месяц квартала с одинаковым средним квадратическим отклонением. Если информация охватывает период $[t_1, t_2]$, то для прогноза себестоимости на следующий квартал (на $(t+1)$, $(t+2)$, $(t+3)$ -й месяцы) достаточно только основного уравнения (4.5). Затем, зная центры прогнозируемых интервалов $\bar{x}_1(t_2+1)$, $\bar{x}_1(t_2+2)$, $\bar{x}_1(t_2+3)$, по уравнениям (4.6) найдем значения $\bar{x}_4(t)$, $\bar{x}_5(t)$, $\bar{x}_{13}(t)$ в те же моменты времени. Теперь будем иметь полную информацию уже за больший период — $[t_1, t_2+3]$. Используя ее как начальную, можно найти предполагаемые значения себестоимости в моменты t_2+4 , t_2+5 , t_2+6 и т. д.

По динамической модели себестоимости были найдены центры прогнозируемых интервалов значений себестоимости для каждого месяца следующего года и сопоставлены с фактическими данными.

Таблица 4.5

Сопоставление прогнозируемых и фактических значений

Месяц	Прогнозируемое значение себестоимости	Фактическое значение себестоимости	Отклонение	
			абсолютное	в % к фактическому значению
I	8,98	9,05	-0,07	0,77
II	8,98	9,14	-0,16	1,75
III	8,89	8,94	-0,05	0,56
IV	8,94	8,90	0,05	0,56
V	8,97	8,93	0,04	0,45
VI	8,84	8,90	-0,06	0,67
VII	8,92	8,82	0,10	1,13
VIII	8,87	8,79	0,08	0,91
IX	8,89	8,87	0,02	0,23
X	8,93	8,96	-0,03	0,33
XI	8,91	8,98	-0,07	0,78
XII	8,92	9,05	-0,13	1,44

Результаты расчетов, приведенные в табл. 4.5, показывают, что прогноз достаточно точен.

4.15. Адаптационные динамические модели для стационарных производственных процессов применимы к более низкому, чем предприятие, уровню организации промышленного производства — технологическим агрегатам. Основное назначение таких моделей — внутриводской технико-экономический анализ и управление производством. Показатели производственного процесса на технологическом агрегате часто значительно варьи-

руют во времени даже при стабильности основных производственных условий. Поэтому при исследовании многих производственных процессов параметры регрессии изменяются во времени. Чтобы моделировать такое изменение, необходимо знать свойства этого процесса (стационарность, корреляционные функции и т. д.). Отсутствие подобных знаний приводит к использованию информации по ходу ее поступления от объекта. Такой путь построения модели называют адаптационным⁴.

Под адаптацией понимается «процесс изменения параметров и структуры системы, а возможно, и управляющих воздействий на основе текущей информации с целью достижения определенного, обычно оптимального состояния системы при начальной неопределенности и изменяющихся условиях работы» (Цыпкин, 1968, с. 58). Адаптационные модели весьма эффективны при моделировании производственных процессов на технологических агрегатах. На более крупных промышленных объектах (цехи, предприятия, комбинаты) четко прослеживаются тренды, и в моделях таких объектов анализом охватываются более длительные промежутки времени.

Следует различать две принципиально различных процедуры адаптации моделей: в режиме реального времени и методом «скользящего интервала», когда, по существу, реализуется идея дискретно-непрерывного моделирования для отдельного объекта. В практике экономических исследований наибольшее значение имеет вторая процедура. Она предполагает, во-первых, выбор интервала времени, в течение которого можно считать выявленную зависимость стабильной, и определение точки изменения характера зависимости; во-вторых, определение необходимого количества наблюдений на этом интервале времени; в-третьих, обоснование полученной зависимости по данным текущей информации.

Для решения первой задачи могут быть использованы методы распознавания образов, дисперсионного анализа (Шеффе, 1963) и их модификации для данного случая (Mac, Gre, Cartetion, 1970). Подходы к решению второй задачи представлены, например, в работах В. Т. Кулик (1968) и В. В. Матушевского (1965).

⁴ Проблемы и методы осуществления адаптации обсуждаются в ряде работ советских и зарубежных авторов (Литвинов, 1966; Райбман, Чадеев, 1966; Цыпкин, 1968; Albert, Sittler, 1965; и др.).

Для решения третьей задачи организуется процесс обновления регрессионной зависимости. Коррекцию модели можно проводить не после каждого нового наблюдения, а по накоплении k новых наблюдений. Тогда отбрасываются k первых и прибавляются k новых наблюдений. Иногда первые наблюдения сохраняют и прибавляют к ним новые. При адаптации по способу «скользящего интервала» необходимы значительные пересчеты с запоминанием больших объемов информации. Упрощает процедуру использование итеративных методов пересчета, предложенных в ряде работ (Кулик, 1964; Маршак, 1970).

В некоторых ситуациях рационально для адаптации сочетать методы пассивного и активного эксперимента, в частности методы регрессионного анализа и эволюционного планирования (Экономико-статистические исследования, 1969). Построение регрессионной модели по отчетным материалам позволяет выявить наиболее существенные факторы процесса и направления их влияния на технико-экономические показатели. Затем методами эволюционного планирования учитывается дрейф процесса и уточняется влияние выделенных существенных факторов.

§ 2. Построение динамических моделей совокупности объектов (на базе частных статических ЭСМ)

4.16. Общая динамическая модель может быть синтезирована из частных статических моделей, если исходная информация охватывает сравнительно большое количество объектов, а протяженность динамических рядов невелика, и из частных динамических — при достаточной большой протяженности временных рядов.

Если совокупность включает небольшое число объектов и сравнительно короткие временные реализации, то используют обычно прием «заводо-годы».

В этом параграфе рассматривается наиболее типичная для практики экономико-статистических исследований первая ситуация (при $N \geq 30$, $t \leq 10$).

4.17. Построение общей динамической модели на базе частных статических показано на схеме 3. Оно основа-

но на комбинации методов анализа пространственно-временной однородности совокупности (операции 1—4, 6) и методов динамизации частных статических моделей (операции 5, 7, 8). Большое внимание уделяется проблеме учета характера топологической структуры совокупности, так как ее изменение во времени следует считать одной из основных причин нестабильности параметров частных статических (пространственных) моделей.

Приведенную схему можно анализировать в двух направлениях: по вертикали (последовательность этапов) и по горизонтали (пути построения моделей различных типов). В зависимости от характера исходной типологической структуры совокупности и поведения ее элементов во времени выделяется пять путей построения общей динамической модели (A, B, C, D, F).

4.18. Наиболее общий случай — построение модели типа A — динамической дискретно-непрерывной модели с изменяющимися во времени параметрами. Для линейной аппроксимации ее можно записать следующим образом:

$$Y_l(t) = \sum_{j=1}^p \sum_{\tau=1}^{t_j} a_{j\tau}^l(t) X_j(t - \tau) + \sum_{\tau=1}^{t_0} b_{\tau}^l(t) Y(t - \tau) + c^l(t);$$

$$\vec{X}_l \in A_l; l = 1, 2, \dots, S,$$

где S — число классов;

p — число независимых переменных;

t_j, t_0 — число временных сдвигов для x_j, y ($j=1, 2, \dots, p$).

Модели различных типов применяются в разных ситуациях: типа A — когда имеется пространственно-временная неоднородность совокупности; типа B — выявлена пространственная структура, но она устойчива во времени; типа C — совокупность представляет собой один пространственный класс, но его характеристики существенно изменяются во времени; типа D — совокупность однородна в пространстве и во времени; типа F — нет достаточной информации для построения статистической модели.

Рассмотрим содержание отдельных этапов схемы.

4.19. Этап 1. Анализ пространственной однородности по методике выполнения аналогичен первому этапу схемы 3, но проводится T раз — для каждого года отдельно. В результате получаем характеристики классов: (x_i^l, \sum_i^l) — вектор средних значений и корреляционную матрицу для

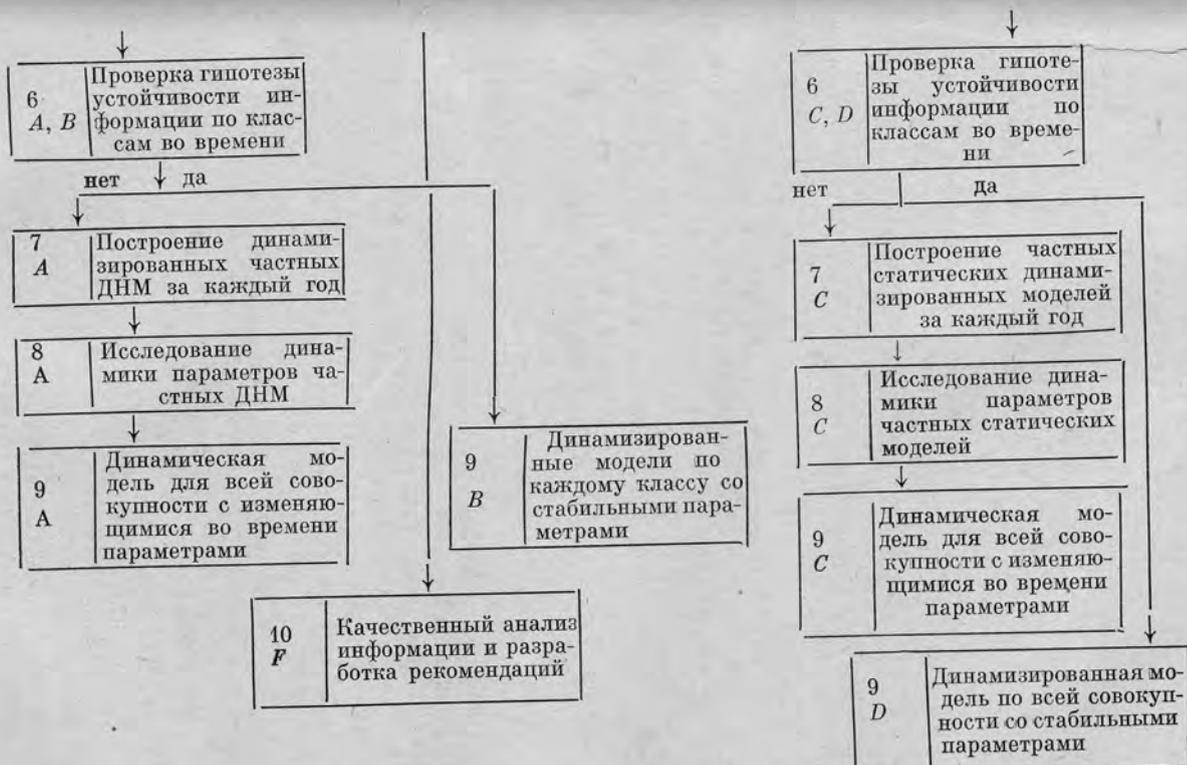
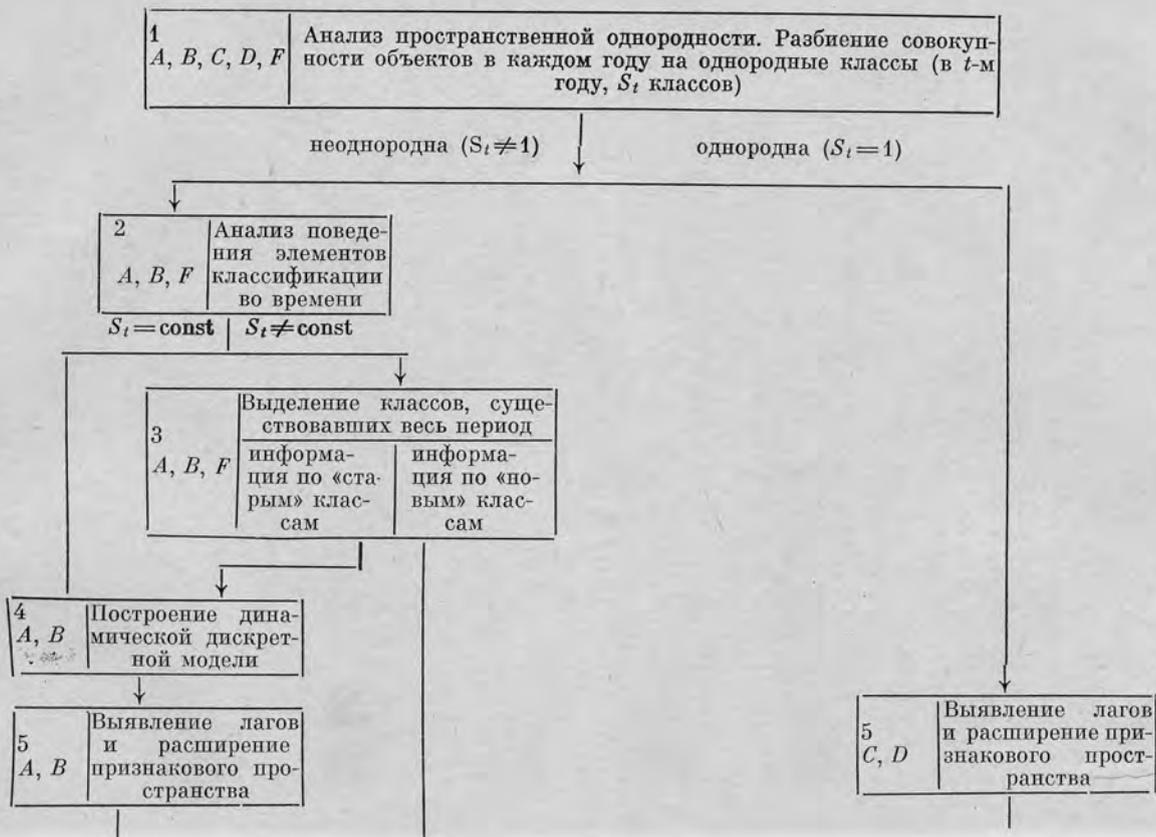


Схема 3. Построение общей динамической модели на базе частных статических.

каждого года (t) и класса (l); n_l^t — количество объектов в каждом классе и S_t — число классов в t -м году. Если число классов в каждом году равно единице ($S_t = 1$), то совокупность считается пространственно однородной и осуществляется переход к пятому этапу, минуя этапы анализа классификаций во времени.

Если совокупность пространственно неоднородна ($S_t \neq 1$), переходим к этапу 2.

4.20. **Этап 2. Анализ поведения элементов пространственной классификации** во времени — один из наиболее важных этапов, принципиально отличающих схему построения динамической ЭСМ от статической. Именно развитие, нестабильность типологической структуры определяют основные сложности при построении динамических моделей.

Изменение типологической структуры происходит за счет двух источников — развития действующих в базисный период предприятий и ввода в эксплуатацию новых предприятий и включает следующие элементарные процессы:

дрейф характеристик классов в исследуемом периоде; изменение численности классов во времени за счет перемещения объектов из класса в класс и появления новых объектов;

появление новых классов и исчезновение отдельных старых.

Первые два элементарных процесса носят систематический, плавный характер, но скорость их протекания может быть переменной. Третий — изменение набора классов — происходит скачкообразно.

Влияние развития действующих в базисный период предприятий на типологическую структуру совокупности зависит от того, на каком участке траектории своего развития находятся предприятия. На большей части этой траектории классовая принадлежность объекта не меняется. Однако в отдельные периоды происходит резкое изменение его состояния. Сюда относится реконструкция, после которой предприятие может перейти в другой из существующих классов и даже образовать новый класс; период освоения производственной мощности; радикальные изменения технологии, характера сырья и выпускаемой готовой продукции и пр. В результате происходит перераспределение объектов между классами, но набор классов, как правило, остается стабильным в течение сравнительно боль-

шого промежутка времени, так как радикальные изменения способа производства в развитии предприятия сравнительно редки.

Ввод в эксплуатацию новых предприятий всегда изменяет объем совокупности и элементы ее структуры. За счет новых предприятий увеличивается удельный вес прогрессивных классов и могут появиться новые классы.

Главное содержание второго этапа заключается в исследовании изменения количества классов во времени и их пообъектного состава с помощью графика переходов объектов из класса в класс. Если количество классов стабильно в течение анализируемого периода времени ($S_t = \text{const}$), то переходим к четвертому этапу⁵. Если в течение исследуемого периода имеет место нестабильность числа классов (появляются «новые» классы, исчезают «старые»), то возникает необходимость в третьем этапе.

4.21. **Этап 3. Выделение классов, существующих длительный период времени** (устойчивых классов), предполагает анализ пространственных классификаций и графика переходов. Выделив устойчивые в течение некоторого периода времени «ядра» классов, формируем новую информацию для этой части объектов и переходим к четвертому этапу — построению динамической ЭСМ. Для оставшихся «новых» неустойчивых классов модель может быть построена лишь по мере накопления информации. Имеющаяся по таким классам информация может служить для проведения качественного анализа и разработки некоторых рекомендаций (этап 10).

Таким образом, коротко содержание второго и третьего этапов может быть сформулировано как установление соответствия между классами в разные годы и выделение устойчивых классов. Особую роль в установлении соответствия между классами играет матрица связи между ними (обозначим ее Λ)

$$\Lambda = \|\lambda_{\tau t}\|, \tau, t = 1, \dots, T,$$

где $\lambda_{\tau t}$ есть мера связи классификации t -го и τ -го годов.

⁵ Опыт многочисленных исследований показывает, что стабильное число классов в сложившихся отраслях обычно сохраняется в течение 7—10 лет и, следовательно, сфера применения предлагаемого подхода достаточно широка.

В качестве таких мер могут использоваться, например, коэффициенты взаимной сопряженности Пирсона или Чупрова, коэффициент близости двух классификаций и другие аналогичные показатели.

Можно всегда предположить, что $\lambda_{ii} = \max_j \lambda_{ij} = 1$.

Следовательно, $\lambda_{ij} \leq 1$, $i, j = 1, \dots, T$. Для получения критерия устойчивости (преемственности) классификаций по годам удобно перестроить матрицу Λ по какому-либо алгоритму диагонализации, например по алгоритмам, предложенным А. В. Беккером и М. Л. Лукацкой (Вопросы экономико-статистического моделирования, 1970, с. 273—284), В. Я. Лумельским (1970) и др.

Разумеется, принятые решения о соответствии классов разных лет (т. е. о том, что тот или иной класс $(t+1)$ -го года есть развитие определенного класса t -го года) связывается с некоторым порогом, который задается в зависимости от содержания конкретных задач.

4.21. Этап 3. Выделение классов, существующих длительно удобно начинать с проверки гипотезы (H_0) об идентичности пространственных классификаций для всех годовых выборок (Андерсон, 1963; Верховская, 1971; Проблемы экономико-статистического анализа, 1969, с. 25—39).

Если нулевая гипотеза об устойчивости классификаций во времени отвергается, возникает задача учета «дрейфа» характеристик классов $-\bar{x}_i^t, \sum_i^t$. Это усредненные величины (по определенному числу объектов), поэтому их изменения во времени обычно носят плавный, монотонный характер и достаточно хорошо аппроксимируются простыми функциями времени.

4.23. Этап 5. Выявление лагов, расширение признакового пространства осуществляется с помощью общепринятых в математической статистике методов построения и анализа взаимокорреляционных функций (Тинтнер, 1965; Хеннан, 1966).

Выявив лаги для получения расширенного признакового пространства, либо заменяем соответствующие столбцы аналогичными, но смещенными на величину лага, либо дополняем таблицу исходных данных этими смещенными столбцами. Кроме того, вводим новые переменные, связанные с отсчетом времени, — чаще всего t и t^2 . О содержательном смысле этих аргументов и об их интерпретации речь шла в § 1 данной главы. Расширение признакового

пространства проводится при моделировании однородной совокупности объектов (операция 5СД). Операция 5АВ выполняется для каждого класса отдельно.

4.24. Этап 6. Проверка гипотезы устойчивости внутриклассовых взаимосвязей во времени отличается от проверки нулевой гипотезы на четвертом этапе тем, что здесь акцент делается на анализ устойчивости внутриклассовых связей. При этом используется расширенная информация, полученная на предыдущем этапе. Алгоритмы проверки одни и те же. Если гипотеза об устойчивости принимается, то строится динамизированная модель со стабильными во времени параметрами (метод «заводо-годы») по каждому классу (9В) или по всей совокупности объектов (9Д); если гипотеза отвергается, т. е. внутриклассовые взаимосвязи изменяются во времени, то переходим к этапу 7.

4.25. Этап 7. Построение динамизированных статистических моделей для каждого года предполагает введение в статистическую ЭСМ некоторых элементов динамики: учета лагов и фактора времени. Такие модели строятся для каждого года по расширенной информации в соответствии с методикой построения, изложенной в гл. 3. Для однородной совокупности (7С) это динамизированные непрерывные модели, для неоднородной (7А) — динамизированные дискретно-непрерывные модели.

4.26. Названные модели могут быть использованы и для проверки нулевой гипотезы на этапе 6. В принципе можно изменить последовательность этих двух этапов: сначала построить модели, затем проверять H_0 .

В данном случае информация годовой выборки заменяется ее кратким описанием — частным уравнением регрессии. Понятно, что это упрощает путь к нашей цели — построению динамической модели для пространственно-однородного класса, но в то же время может огрубить результаты проверки гипотезы устойчивости за счет ошибок аппроксимации. Проверка нулевой гипотезы включает анализ: устойчивости структуры пространственных моделей (т. е. набора информативных переменных) и их совокупного влияния на моделируемый показатель; соотношений в степени влияния отдельных факторов; значений свободного члена и коэффициентов регрессии.

Способы решения двух первых задач сравнительно просты. Для каждого из пространственных уравнений регрессий независимо определяется набор информативных пере-

менных и рассчитываются коэффициенты множественной корреляции. Полное (или близкое) совпадение информативных наборов позволяет сделать вывод о временной устойчивости структуры моделей, а перекрытие доверительных интервалов множественных коэффициентов корреляции свидетельствует об устойчивости совокупного влияния информативных признаков на моделируемый показатель.

Один из наиболее простых приемов проверки устойчивости соотношения величины коэффициентов регрессии при отдельных переменных заключается в ранжировке переменных по степени влияния на функцию. Постоянная величина рангов отдельных переменных или слабое изменение их во времени свидетельствуют об определенной устойчивости соотношений коэффициентов регрессий. Приведем пример ранжировки переменных модели одного из исследований по лесозаготовительному производству (табл. 4.6). Из приведенных данных видно, что по соотно-

Таблица 4.6

Ранги факторов (переменных модели) по уровню β -коэффициентов

Фактор	Первый год	Второй год	Третий год	Четвертый год
Объем вывозки	1	1	1	1
Состав насаждений:				
сосна	3	3—4	3	3
береза-осина	5	6	—	—
Тип примыкания складов	4	3—4	4	4—5
Стоимость производственных фондов	2	2	2	2
Среднее расстояние вывозки	6	5	5	4—5

шению в силе влияния факторов сравниваемые модели устойчивы во времени.

4.27. Для проверки устойчивости во времени значений параметров (a не их соотношений) можно использовать формальные методы проверки идентичности двух или нескольких регрессий. Проверяются следующие гипотезы относительно сравнительной величины коэффициентов регрессий ($\hat{\sigma}$) и свободного члена (a)⁶:

⁶ Индексы (1) и (2) означают принадлежность к первой и второй выборке соответственно.

1. $a^{(1)} = a^{(2)}$, $\hat{\sigma}^{(1)} = \hat{\sigma}^{(2)}$ — регрессии идентичны по сдвигу и наклону;

2. $a^{(1)} \neq a^{(2)}$, $\hat{\sigma}^{(1)} = \hat{\sigma}^{(2)}$ — регрессии различаются только по сдвигу;

3. $a^{(1)} = a^{(2)}$, $\hat{\sigma}^{(1)} \neq \hat{\sigma}^{(2)}$ — регрессии различаются по углу наклона;

4. $a^{(1)} \neq a^{(2)}$, $\hat{\sigma}^{(1)} \neq \hat{\sigma}^{(2)}$ — регрессии различны по сдвигу и наклону.

Гипотеза об идентичности двух уравнений регрессии может проверяться методом, применяемым к общим линейным гипотезам (Chow, 1960). Проверяется нулевая гипотеза $H_0: \beta = \beta_1 = \beta_2$ (регрессия приведена к стандартному виду) против альтернативы $H_a: \beta_1 \neq \beta_2$.

Модель общей линейной гипотезы может быть записана следующим образом:

$$y_1 = x_1\beta_1 + 0\beta_2 + \varepsilon_1;$$

$$y_2 = 0\beta_1 + x_2\beta_2 + \varepsilon_2.$$

При верной нулевой гипотезе сумма квадратов отклонений равна сумме квадратов отклонений при альтернативной гипотезе плюс сумма квадратов отклонений двух наборов оценок y для этих же гипотез. Отношение двух последних сумм, деленное на соответствующее число степеней свободы, как показал Chow, следует F -распределению при верной нулевой гипотезе. Поэтому вычисляется

$$F = \frac{S_1(n+m-p) - [S_2/(n-p) + S_3/(m-p)]}{S_2/(n-p) + S_3/(m-p)} = \frac{S_5/p}{S_4/(n+m-2p)},$$

где S_1 — сумма квадратов отклонений общей регрессии, рассчитанной по объединенной совокупности $m+n$ наблюдений;

S_2 — сумма квадратов отклонений регрессии, рассчитанной по n наблюдениям;

S_3 — сумма квадратов отклонений регрессии, рассчитанной по m наблюдениям;

p — число оцениваемых параметров регрессии.

Полученное значение F сравнивается с табличным F_α , которое находится для соответствующих степеней свободы p и $m+n-2p$ и уровня значимости α . Если $F > F_\alpha$, то ну-

левая гипотеза об идентичности уравнений регрессии для двух групп наблюдений отвергается.

Рассмотренный метод имеет тот недостаток, что в случае, если нулевая гипотеза отвергается, то неизвестно, чем обусловлено различие уравнений регрессии — их сдвигом или различием коэффициентов регрессии и каких именно.

Метод фиктивных переменных (Gujarati, 1970) свободен от указанного выше недостатка и состоит в построении уравнения регрессии по объединенной совокупности двух выборок. Общий вид регрессионного уравнения для объединенной совокупности следующий:

$$y = a_0 + Da_1 + \sum_{j=1}^p x_j c_j + \sum_{j=1}^p D x_j b_j \quad (4.7)$$

($D = 1$, если наблюдение принадлежит первой совокупности и $D = 0$ — для второй). Если a_1 статистически значим, то величина сдвига для первой выборки равна $a_0 + a_1$, для второй — a_0 . Если a_1 статистически незначим, то a_0 дает оценку общего для обеих совокупностей сдвига. Если b_j статистически значим, значение соответствующего коэффициента для первой выборки $b_j + c_j$, для второй — c_j . Если b_j статистически незначим, то c_j дает оценку общего для обеих выборок значения коэффициентов.

При помощи метода фиктивных переменных можно установить, по каким признакам различаются две регрессионные модели. Зная причину различий в уравнениях регрессии по годам, можно выявить определенные закономерности и использовать их при построении динамической модели.

Для проверки устойчивости параметров регрессии во времени могут быть также использованы аппарат теории информации (Кульбак, 1967) и методы дисперсионного анализа (Шеффе, 1963).

Материалы целого ряда прикладных исследований позволяют сделать следующие заключения о наиболее типичном поведении элементов пространственных моделей во времени. Из параметров регрессии наименее устойчив свободный член, он меняется от года к году. Более устойчивы коэффициенты регрессии, которые обычно для двух последующих лет различаются несущественно. Еще устойчивее во времени набор существенных (информативных) факторов, их ранжировка по степени влияния на моделируемый показатель и коэффициенты множественной кор-

реляции. Как правило, эти характеристики пространственных моделей остаются стабильными в течение 5—8 и более лет.

4.28. Проверка однородности статистических данных при регрессионном анализе лесозаготовительного производства показала, что по критерию Chow уравнения можно считать идентичными только по парам последующих лет⁷.

Величины средних квадратических отклонений σ_{b_j} и t -отношений — b_j/σ_{b_j} (t — фактических) для коэффициентов при фиктивных переменных в регрессионном уравнении (4.7) при проверке рассмотренным выше методом для тех же данных за два года оказались следующими:

	σ_{b_j}	b_j/σ_{b_j}
Удельный вес сосны	0,71	0,03
Удельный вес березы	0,82	1,41
Средний запас древесины на 1 га	0,57	0,95
Среднее расстояние вывозки	1,19	0,86
Объем вывозки	2,76	0,006
Удельный вес вывозки:		
автомашинами	0,56	0,36
тракторами	0,86	0,54
Свободный член регрессии	8,5	19,7

Из сравнения t фактического с t табличным следует, что статистически значимо изменение свободного члена, а различия коэффициентов при переменных статистически незначимы.

4.29. Этап 8. Исследование динамики параметров частных моделей включает конструирование временных рядов для каждого параметра динамизированных моделей, выявление вида соответствующих тенденций (см. § 1 настоящей главы), запись структуры общей динамической модели. Смысл рассмотренных операций заключается в обосновании вида общей динамической модели (см. п. 4.18). После этого оцениваются численные значения ее параметров с помощью одного из известных алгоритмов. Для линейного случая методика расчета рассмотрена в работе А. А. Френкеля (1972).

⁷ Подробно эта проверка проведена в работе М. Л. Луцаккой, Т. Р. Сазоновой, Т. В. Филипповой «О некоторых методах сравнения регрессий, построенных по разным выборкам» (Вопросы построения и применения статистических моделей экономических показателей предприятий, ч. II. Новосибирск, 1971, с. 127—137).

§ 3. Построение общей динамической модели совокупности объектов (на базе частных динамических ЭСМ)

4.30. В общем случае указанный в заголовке процесс включает: а) построение для каждого объекта частной динамической модели на основе анализа многомерных временных рядов (см. § 1 настоящей главы); б) многомерную классификацию этих моделей по их параметрам и в) конструирование для каждого сформированного класса внутриклассовой динамической модели. Если число объектов значительно и частные динамические модели включают сравнительно много переменных, то реализация схемы построения общей динамической модели весьма трудоемка.

4.31. В нашей практике большее распространение получила модификация общей схемы, основанная на синтезе идей пространственной многомерной классификации и анализа одномерных временных рядов.

I этап — формирование множества реализаций экономического показателя. Во многих случаях здесь нет необходимости рассматривать непрерывные функции, можно ограничиться заданием индивидуальных кривых в виде набора дискретных точек в пространстве (Y, T) . Чтобы обеспечить сопоставимость, на кривых динамики показателей выделяются характерные точки и участки. Под характерными точками понимаются моменты существенного изменения структурных свойств обрабатываемой реализации, характерные участки представляют собой отрезки кривых, заключенные между соседними характерными точками. Для выделения характерных точек можно использовать содержательный, профессиональный анализ и формальные приемы, например введение функционала отличия и анализ его поведения во времени. Частный случай такого функционала — ошибка фильтрации (следящий сигнал) при выравнивании ряда методами скользящего среднего или экспоненциального сглаживания. Появление локальных экстремумов и превышение заранее заданного порога фиксирует наличие характерных точек. В развитии предприятия характерные точки — это обычно момент пуска в эксплуатацию, моменты полного освоения проектной мощности и проектных значений экономических показателей, проведения реконструкции или внедрения новых технологических способов производства и т. д. Обеспе-

ние сопоставимости может потребовать приведения отрезков к единой длине путем перехода на условные измерители. Итак, при формировании ансамбля временных реализаций экономических показателей необходимо решить две задачи:

расчленив индивидуальные временные ряды показателей на отдельные характерные участки с определенными структурными свойствами;

обеспечить сопоставимость этих отрезков путем приведения к единой длине (например, можно принять длину каждого индивидуального отрезка за 100% или за 1).

4.32. II этап — формирование обучающей выборки в пространстве выходных показателей (\vec{Y}, T) , для чего осуществляется многомерная классификация объектов по характеру кривых (вернее, анализируемых отрезков кривых). При классификации индивидуальных отрезков кривых могут использоваться различные алгоритмы таксономии, а отдельные временные реализации задаваться как всеми точками, так и наиболее информативными. Результат этого этапа — построение алфавита классов типичных кривых для определенного характерного участка траектории развития предприятия (например, для периода освоения) и выбор предприятий-представителей каждого класса.

4.33. III этап — математическое описание полученной на II этапе системы класс-эталонов в новом признаковом пространстве, т. е. переход из пространства «показатели-время» (\vec{Y}, T) в пространство «факторы производства-время» (\vec{X}, T) и обучение распознаванию классовой принадлежности объектов в нем.

4.34. IV этап — преобразование по результатам обучения пространства (\vec{X}, T) в пространство (\vec{X}', T) .

4.35. V этап — формирование классов в преобразованном пространстве (\vec{X}', T) методами таксономии. Классификация, полученная на этом этапе, может быть названа «ковечной», она выступает как дискретная часть динамической модели показателя.

4.36. VI этап — построение типичных кривых динамики показателя для всех реализаций, попавших в данный класс. Содержание этапа заключается в поиске наиболее подходящего способа описания внутриклассовой динамики

показателя, заданного конечным числом реализаций. Наиболее часто для решения этой задачи используются методы регрессионного анализа, т. е. тенденция показателя отыскивается как явная функция времени. Иногда дополнительно используются авторегрессионные преобразования.

В общем виде дискретно-непрерывная модель вышеописанного типа может быть записана следующим образом:

$$Y = \begin{cases} f_1[t, Y(t-1), Y(t-2), \dots, Y(t-k)\vec{X}] & \text{при } \vec{X} \in A_1 \\ \dots \\ f_j[t, Y(t-1), Y(t-2), \dots, Y(t-k)\vec{X}] & \text{при } \vec{X} \in A_j \\ \dots \\ f_m[t, Y(t-1), Y(t-2), \dots, Y(t-k)\vec{X}] & \text{при } \vec{X} \in A_m. \end{cases}$$

Чтобы обеспечить возможность прогнозирования, набор факторов производства \vec{X} обычно строят таким образом, чтобы их значения были известны до начала периода прогнозирования, т. е. \vec{X} формируется из признаков со значительными лагами. Пример реализации указанного подхода рассмотрен в гл. 6, § 2.

§ 4. Построение общей динамической модели по исходной информации малой размерности

4.37. В экономической практике часто возникает необходимость построить и проанализировать динамическую модель (с возможностью использования ее в качестве прогнозной) для небольшого числа промышленных объектов. В этом случае обычно применяют так называемый прием «заводо-годы», при котором анализируется одна общая совокупность без дифференциации во времени. Каждый объект при этом выступает в качестве самостоятельного наблюдения несколько раз, увеличивая объем совокупности.

Такой прием не всегда правомерен, так как исследуется фактически не одна совокупность, а смесь нескольких совокупностей с характеристиками и взаимозависимостями, изменяющимися во времени. В каждом отдельном случае необходимо либо показать, что полученную совокупность заводов-лет можно без особого ущерба считать однородной (1), либо выделить действительно однородные совокупности устойчивых во времени предприятий и для них

строить модели (2), либо применить определенные преобразования информации, обеспечивающие повышение однородности совокупности (3).

4.38. Если идти по второму пути, применяя методы теории распознавания образов (Розин, Лукацкая, Овчинникова, 1975), то такое выделение однородных во времени классов проводится по информации за весь период времени. Уменьшая количество объектов в классе, можно достичь некоторой однородности во времени, а применяя затем к каждой группе прием «заводо-годы», — восполнить объем информации, необходимый для построения внутриклассовых непрерывных моделей.

4.39. Существует прием, при котором однородность во времени достигается за счет уменьшения длительности периода. Весь период T разбивается на S интервалов; выдвигается и проверяется гипотеза о том, что за время, равное длине интервала (T/S), статистические характеристики и взаимозависимости остаются постоянными или изменяются несущественно. По информации о каждом из интервалов строятся статистические модели. Для каждого параметра модели есть свой временной ряд длиной S . Анализ этих рядов позволит сконструировать общую динамическую модель (см. § 2 настоящей главы). Однако использовать такой прием можно лишь при значительной длине временных рядов, а подготовка подобной информации связана с большими трудностями.

4.40. Если идти по третьему пути, то для преобразования информации можно использовать переход к относительно величинам

$$z_t^{(k)} = y_t^k / \bar{y}_t^{(k)}, \quad u_t^{(l)} = x_t^{(l)} / \bar{x}_t^{(l)} \quad (k = 1, \dots, m; \\ l = 1, \dots, p; t = 1, \dots, T), \quad (4.8)$$

где $y_t^{(k)}$, $x_t^{(l)}$ — исходные данные (см. табл. 4.1).

Дисперсии и математические ожидания случайных величин $z_t^{(k)}$, $u_t^{(l)}$ будут равны соответственно

$$D(z_t^{(k)}) = D(y_t^{(k)}) / (\bar{y}_t^{(k)})^2; \quad D(u_t^{(l)}) = D(x_t^{(l)}) / (\bar{x}_t^{(l)})^2; \\ M(z_t^{(k)}) \equiv M(u_t^{(l)}) \equiv 1, \quad (k = 1, \dots, m; l = 1, \dots, p; \\ t = 1, \dots, T),$$

а последовательность корреляционных матриц $\{R_t\}$ остается прежней.

Справедливо следующее утверждение: если первоначальная последовательность случайных величин

- а) имеет нормальное распределение;
 б) ее корреляционные матрицы R_t несущественно различаются для разных t ;

в) дисперсии пропорциональны квадратам соответствующих математических ожиданий, то распределение введенных относительных величин

$$\{z_t^{(1)}, \dots, z_t^{(m)}, u_t^{(1)}, \dots, u_t^{(p)}\}, t = 1, \dots, T$$

не изменяется с изменением t .

Таким образом, к новой, преобразованной информации вполне обоснованно применение приема «заводо-годы».

4.41. Построение динамической модели на основе такого преобразования включает следующие этапы.

1. На основании исходной информации о значениях показателей и факторов производства для n предприятий в T моментов времени ($y_{it}, x_{it}^{(1)}, \dots, x_{it}^{(m)}, i=1, 2, \dots, N; t=1, 2, \dots, T$) выбирают некоторое базисное предприятие⁸, характеризующееся в каждом году значениями показателя и признаков, равными

$$y_{\delta t}, x_{\delta t}^{(1)}, \dots, x_{\delta t}^{(m)}, t = 1, 2, \dots, T.$$

2. Выполняют преобразование исходной информации, вводя относительные переменные вида

$$z = y/y_{\delta}, u^{(1)} = x^{(1)}/x_{\delta}^{(1)}, \dots, u^{(m)} = x^{(m)}/x_{\delta}^{(m)},$$

и строят регрессионную модель в новых переменных:

$$z = \sum_{j=1}^m a_j u^{(j)} + a.$$

3. Переходят к первоначальным переменным, для чего записывают модель показателя в виде линейной комбинации признаков с изменяющимися во времени коэффициентами

$$y_t = \sum_{j=1}^m b_{jt} x^{(j)} + b_t, t = 1, 2, \dots, T,$$

где

$$b_{jt} = y_{\delta t} \cdot a_j / x_{\delta t}^{(j)}, b_t = y_{\delta t} \cdot a, j = 1, 2, \dots, m,$$

⁸ В качестве базисного предприятия можно взять гипотетическое «среднее» предприятие — среднее в статистическом смысле.

и пересчитывают статистические оценки для полученной модели.

4. Аппроксимируют изменение коэффициентов b_t, b_{jt} ($j=1, \dots, m$) во времени достаточно простыми функциями и записывают модель в виде

$$y(t) = \sum_{j=1}^m b_j(t) x^{(j)} + b(t).$$

4.42. Приведем пример построения динамической модели по предложенной схеме. Исходной информацией служат полные затраты и условный объем производства для девяти цементных заводов за пятилетний период (в условных единицах):

Номер завода	Полные затраты (y) по годам				
	1	2	3	4	5
1	10,639	10,332	13,977	18,457	17,344
2	3,556	4,036	4,235	4,685	4,424
3	6,476	7,144	7,581	6,418	5 619
4	10,675	12,189	12,240	12,461	10,309
5	10,640	10,330	11,266	11,758	10,752
6	7,508	7,991	7,963	7,823	7,256
7	11,280	11,704	12,494	15,459	13,583
8	8,754	9,606	10,631	10,962	9,114
9	4,120	4,076	4,362	4,450	4,026

Номер завода	Объем производства (x) по годам				
	1	2	3	4	5
1	7,14	8,24	10,75	14,20	16,82
2	2,85	3,95	4,33	4,69	4,80
3	5,58	6,19	6,84	6,60	6,91
4	6,23	7,71	8,25	8,33	8,23
5	5,54	4,22	5,42	6,29	6,82
6	3,93	4,31	4,62	4,72	5,19
7	7,66	7,16	7,82	9,62	10,19
8	4,68	4,12	5,56	5,43	5,35
9	0,57	0,65	0,74	0,74	0,77

(Полужирные цифры — для завода, принятого за базисный.)

1. Анализ исходных данных показывает, что небольшой цементный завод (№ 9) не совсем типичное предприятие, так как мощность его в 5—6 раз меньше следующего за ним по мощности завода (№ 2), а затраты примерно на том же уровне. Поэтому в качестве базисного берем завод № 2 и переходим к относительным величинам $z = y/y_2, u = x/x_2$.

2. Значения первоначальной дисперсии относительного показателя затрат (z) и остаточных дисперсий при аппроксимации (методом наименьших квадратов) этих затрат некоторыми простыми функциями u для каждого года отдельно и по всему массиву приведены в табл. 4.7.

Так как значения дисперсии для каждого года и всего массива получены по существенно разному числу наблюдений, причем

Таблица 4.7

Статистика	Первый год	Второй год	Третий год	Четвертый год	Пятый год	Весь массив
Первоначальная дисперсия	73,53	53,76	56,35	58,53	57,34	55,06
Остаточные дисперсии для функций:						
$au + b$	0,698	0,741	0,589	0,673	0,763	0,937
$au^2 + bu + c$	0,145	0,229	0,189	0,193	0,164	0,793
$a \ln^2 u + b \ln u + c$	1,508	1,686	1,427	2,108	2,609	2,058
$au^3 + bu^2 + cu + d$	0,145	0,223	0,187	0,192	0,158	0,315
$a \ln^3 u + b \ln^2 u + c \ln u + d$	0,200	0,229	0,238	0,278	0,297	0,470

для каждого года их всего 9, то ясно, что количественное сравнение этих значений необоснованно. Однако характер изменения остаточных дисперсий при переходе от одной функции к другой в общем такой же, как и для всего массива. Анализ дисперсий показывает, что в качестве аппроксимирующей функции можно выбрать полином третьей степени.

$$z = -0,0127 u^3 + 0,297 u^2 + 0,234 u + 1,116.$$

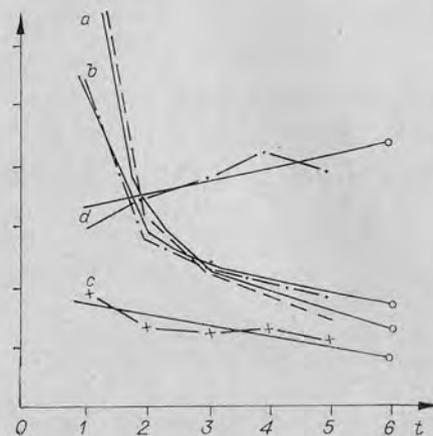
3. Переход к первоначальным переменным позволяет записать это общее уравнение в виде частных для каждого года:

$$\begin{aligned} y_1 &= -0,001950 x^3 + 0,1300 x^2 + 0,292 x + 3,968, \\ y_2 &= -0,000832 x^3 + 0,0768 x^2 + 0,239 x + 4,504, \\ y_3 &= -0,000663 x^3 + 0,0671 x^2 + 0,229 x + 4,726, \\ y_4 &= -0,000577 x^3 + 0,0633 x^2 + 0,234 x + 5,228, \\ y_5 &= -0,000508 x^3 + 0,0570 x^2 + 0,216 x + 4,937. \end{aligned}$$

4. Динамика коэффициентов статистической модели ясно прослеживается на рисунке. Можно считать, что для коэффициентов c и d аппроксимирующими (обозначены сплошными линиями) будут прямолинейные зависимости, для a и b — гиперболические. Записав эти коэффициенты в виде соответствующих функций времени, получим динамическую модель полных затрат

$$y(t) = a(t)x^3 + b(t)x^2 + c(t)x + d(t).$$

Для построенной модели прогноз на шестой год, геометриче-



ски найденный по рисунку, дает следующие значения коэффициентов a, b, c, d : $-0,00044$; $0,054$; $0,185$; $5,35$. Следовательно, по модели вида

$$y_6 = -0,00044 x^3 + 0,054 x^2 + 0,185 x + 5,35$$

рассчитывается прогнозируемая зависимость затрат от объема производства для цементных заводов (в условных единицах).

Раздел II

ВОПРОСЫ ПРИМЕНЕНИЯ ЭКОНОМИКО-СТАТИСТИЧЕСКИХ МОДЕЛЕЙ В АНАЛИЗЕ, НОРМИРОВАНИИ И ПЛАНИРОВАНИИ ПРОМЫШЛЕННОГО ПРОИЗВОДСТВА

Глава 5

ПРИМЕНЕНИЕ ЭСМ В ТЕХНИЧЕСКОМ НОРМИРОВАНИИ, ВНУТРИЗАВОДСКОМ ПЛАНИРОВАНИИ И УПРАВЛЕНИИ

5.1. В некоторых отраслях промышленности вследствие специфики технологического способа производства важно исследовать работу крупных технологических агрегатов. К таким отраслям в первую очередь следует отнести черную и цветную, металлургию, цементную, химическую промышленность и ряд других.

Производственный процесс в них имеет ряд особенностей: производство концентрируется на небольшом числе крупных, дорогостоящих агрегатов, определяющих в основном эффективность работы всего предприятия; преобладают закрытые, аппаратурные физико-химические процессы. Значительный элемент неопределенности в показателях таких процессов обусловлен большим числом факторов, воздействующих на результаты работы агрегатов; наличием значительной неконтролируемой вариации; участием человека в управлении процессами, наконец, невозможностью получения полной информации. В подобных случаях для количественной оценки влияния производственных факторов на показатели работы эффективно применять экономико-статистические методы.

5.2. В настоящей главе из комплекса прикладных применений статистических методов в исследовании работы агрегатов рассматриваются разработка внутризаводских нормативов и выбор экономически выгодных режимов управления агрегатом.

§ 1. Статистическая оценка внутризаводских нормативов

5.3. Для одного и того же показателя разрабатываются различные виды нормативов. Внутризаводские нормативы дифференцируются по области применения (оперативное и технико-экономическое планирование, построение системы оплаты труда рабочих) и сроку действия. Нормативы разного срока действия различаются, во-первых, числом производственных циклов (календарных периодов), средним заданием для которых является соответствующий норматив, и, во-вторых, объемом информации, известной в момент установления норматива. Чем продолжительнее срок действия норматива, тем большее число циклов (периодов) осредняется им, больше факторов может быть учтено по их среднему уровню. Поэтому в принципе нормативы для планирования работы агрегатов на различные сроки не могут быть одинаковыми.

Несмотря на различия в величине и методике расчета, все виды нормативов имеют общие свойства: каждый норматив является плановым (регламентирующим) заданием, основанным на прогнозе возможностей работы агрегата в планируемый период; прогноз базируется на анализе информации о предшествующей работе агрегата и на учете планируемых изменений; во всяком нормировании есть элемент неопределенности, так как при установлении нормативов нет полной информации об условиях будущей работы агрегатов.

5.4. Для построения выделенных типов нормативов используется система переводных коэффициентов. Рассмотрим методику статистической оценки их величины. Технико-экономические показатели работы технологических агрегатов характеризуются не одной нормой, а системой, в которой нами выделяются два элемента — абсолютный уровень норм и их соотношение — так называемые переводные коэффициенты.

Применяемую схему разработки системы норм можно представить следующим образом.

А. Каким-либо способом (на основе теории процесса, экспериментальных исследований и расчетов) строится целевая функция нормирования (зависимость нормируемого показателя от факторов производства) и накладываются ограничения на переменные.

Б. Определяются возможные сочетания факторов, которые для данных условий производства и уровня управления будут заданными $(x_1, x_2, \dots, x_k) - \vec{X}_k$. Они либо регламентированы на более высоких уровнях управления (вид и характер продукции, используемое исходное сырье и топливо и т. д.), либо регулируются по своей природе в пределах рассматриваемого периода времени (технические параметры оборудования и т. д.).

В. Для каждого заданного сочетания задача решается на оптимум по набору факторов автономного управления работой объекта $(x_{k+1}, x_{k+2}, \dots, x_n) - \vec{X}_p$. Их влияние учитывается абсолютным уровнем норм. В условиях исследуемого подразделения исполнитель может изменять факторы в тех или иных пределах.

Г. Определяются переводные коэффициенты — как соотношения расчетных норм.

Группа не поддающихся учету факторов вызывает случайную вариацию показателя, которая не может быть учтена при данных условиях. Поэтому при значительной вариации показателей работы агрегатов нормативы должны быть заданы либо диапазоном значений, либо средней величиной с указанием допустимых отклонений.

5.5. Рассмотренный способ разработки системы норм связан с известными трудностями. Во-первых, нередко число необходимых градаций норм велико и, следовательно, для разработки системы норм потребуется многократно решать оптимальную задачу. Во-вторых, в отдельных случаях построить полную целевую функцию зависимости показателя от всех факторов при имеющихся знаниях о процессе невозможно. Поэтому величина переводных коэффициентов на практике оказывается недостаточно обоснованной, неточной. Существует упрощенный метод построения системы норм, свободный от этих недостатков.

Переменные факторы x_p считаются постоянными и их значения принимаются на уровне фактически сложившихся в данной производственной обстановке. Статистическими методами по данным производственного учета устанавливается величина \bar{Q}_i , соответствующая практически возможным и необходимым сочетаниям $x_k^{(j)}$. Затем одно из сочетаний $x_k^{(0)}$ и соответствующее ему \bar{Q}_0 принимают за условную единицу и рассчитывают систему переводных ко-

эффициентов \bar{Q}_i . Для сочетания условий $x_k^{(0)}$, принятых за базисные, решают задачу на оптимум функции $Q = f(\vec{X}_p)$ и определяют базисную норму N_0 . На основе системы переводных коэффициентов и N_0 восстанавливают всю систему норм.

Таким образом, вместо системы норм можно использовать одну базисную норму и систему переводных коэффициентов. Это удобно для различного рода технико-нормировочных, плановых и других расчетов и имеет самостоятельное значение (так, чтобы рассчитать влияние изменения сортамента, достаточно знать значения переводных коэффициентов, абсолютные значения норм не используются). Кроме того, переводные коэффициенты более устойчивы во времени, чем абсолютные значения экономических показателей. Например, длительность плавки по мере «старения» печи существенно изменяется, а отношения длительностей плавки различных марок практически не зависят от времени.

Сказанное определяет статистическую оценку переводных коэффициентов как важную самостоятельную задачу внутризаводского экономико-статистического исследования.

5.6. Сущность задачи статистической оценки переводных коэффициентов формулируется следующим образом: имея отчетные данные о значении экономического показателя y и о комплексе соответствующих ему неуправляемых условий производства \vec{X}_k , необходимо получить статистическую оценку математического ожидания отношений значений показателя для разных комплексов неуправляемых условий производства.

5.7. Способ решения задачи зависит от вида задания информации в учетной документации. Если исходная информация задана в явном виде, т. е. из материалов учета каждому сочетанию \vec{X}_k может быть поставлено в соответствие определенное фактическое значение экономического показателя, то используется следующая схема расчетов. Определенные условия принимаются за базисные. На основе первичной информации (она обычно представляет собой временной ряд) определяется последовательность индивидуальных значений переводных коэффициентов для каждого сочетания неуправляемых условий производства. Затем исследуется поведение этой последовательности чисел во времени. При устойчивости ряда в качестве норматив-

ной оценки коэффициента может быть принята средняя величина для ряда фактических значений. Если выявляется систематическая тенденция, то строится функция математического ожидания от времени. И наконец, для уменьшения числа коэффициентов агрегируются неуправляемые условия производства.

При неявной форме задания информации (это бывает, когда процесс непрерывный либо циклический с очень короткими циклами) фактические значения экономического показателя обычно учитываются не по сочетаниям неуправляемых условий производства, а по учетным периодам, в течение которых могут иметь место разные комбинации неуправляемых условий производства, и учтенная величина экономического показателя представляет собой какую-то среднюю или накопленную величину. В этом случае дополнительно выявляется величина показателя, соответствующая определенным значениям неуправляемых условий производства.

5.8. Рассмотрим пример оценки переводных коэффициентов производительности нескольких типов машин, занятых выполнением одной и той же работы (Экономико-статистические исследования, 1969). В качестве исходной информации использованы месячные данные о среднечасовой производительности различных типов экскаваторов на угольных карьерах в течение 6 лет. Исследовались два вида работ — вскрыша и добыча на шести карьерах. На основе предварительного анализа выделены неуправляемые условия производства: марка экскаватора, вид работ и наименование карьера.

Обозначим производительность экскаватора основной марки на j -м карьере по s -му виду работ в t -м месяце через Q_{ijt}^s , а производительность экскаватора i -й марки в тех же условиях — Q_{ijt}^s , где Q_0 — основная марка экскаватора. Тогда индивидуальный переводной коэффициент i -й марки экскаватора для соответствующих условий будет равен

$$K_{ijt}^s = Q_{ijt}^s / Q_{0jt}^s \\ (i = \overline{1,6}; s = \overline{1,2}; j = \overline{1,n}; t = \overline{1,72}).$$

Полученные индивидуальные коэффициенты представляют собой временные ряды для добычи и вскрыши соответственно:

$$K_{ijt_1}^{(1)}, K_{ijt_2}^{(1)}, \dots, K_{ijt_{72}}^{(1)}; K_{ijt_1}^{(2)}, K_{ijt_2}^{(2)}, \dots, K_{ijt_{72}}^{(2)}.$$

Проверка их на автокорреляцию всех порядков (при 5%-ном уровне значимости) дает основание принять гипотезу об отсутствии си-

стематической тенденции коэффициентов во времени. При подсчете сезонной волны способом скользящей средней (Юл, Кендалл, 1960) можно было предположить, что весь комплекс сезонных изменений происходит в течение одного года. Ни ряды с трехмесячной скользящей средней, ни месячные индексы не улавливают сезонной волны, поэтому очевидно, что индивидуальные значения переводных коэффициентов распределены во времени случайно и, следовательно, их оценку необходимо искать в виде математического ожидания $K_{ijt}^{(s)}$.

На последнем этапе исследования проверяется возможность агрегирования коэффициентов по карьерам и видам работ, т. е. гипотеза о несущественности различий переводных коэффициентов для экскаваторов данной марки по карьерам. Для экскаваторов ЗШ4-40, занятых на вскрышных работах шести карьеров, проверка методом дисперсионного анализа показала, что выборка неоднородна и единые переводные коэффициенты для всех карьеров не могут быть установлены.

Показатели изменчивости переводных коэффициентов оказались следующими: оценка внутригрупповой дисперсии — 0,062 при числе степеней свободы 295; оценка межгрупповой дисперсии — 1,976 при числе степеней свободы 5. Тогда фактическое значение критерия Фишера составит $F = 1,976/0,062 = 31,77$. Оно намного превышает теоретическое значение этого критерия, которое при вероятности ошибки $P = 0,01$ будет равно $F_{\alpha} = 3,05$.

5.9. Задача определения переводных коэффициентов по видам продукции возникает при циклическом характере производства со сравнительно длительным циклом. Обычно это последовательное производство во времени на одном агрегате различных видов продукции. Для каждого цикла известны неуправляемые условия производства и соответствующая им фактическая величина экономического показателя. Методические затруднения при решении задач такого типа возникают, если с течением времени существенно изменяются состояние и показатели работы агрегатов. При расчете переводных коэффициентов влияние таких динамических сдвигов в работе оборудования должно быть исключено. Можно указать два способа разработки переводных коэффициентов — комбинационную группировку для установления переводных коэффициентов по основным видам продукции и метод множественной регрессии для интерполяции значений на редко выпускаемые виды продукции.

5.10. Рассмотрим методику решения задачи на примере установления переводных коэффициентов длительности плавок различных марок стали.

При использовании метода комбинационной группировки необходимо найти способы элиминирования влияния

всех источников вариации длительности плавки, кроме марки стали. Множество изучаемых плавок на данной печи подразделяется на последовательные группы, внутри которых действием фактора времени можно пренебречь. Исключить влияние индивидуальных и случайных различий — значит определить необходимое число плавок, которое должно быть включено в анализ по каждой марке стали для получения статистически достоверного результата.

Процедура расчета переводных коэффициентов следующая.

1. В качестве основной выбирается марка стали с большим удельным весом в сортаменте, по возможности выплавляемая на протяжении всего исследуемого периода.

2. Для каждого последовательного десятка плавок определяется средняя продолжительность одной плавки. Если обозначить через r номер десятка, \bar{t}_{or} — среднюю продолжительность плавки основной марки стали в r -м десятке, \bar{t}_{ir} — среднюю длительность плавки i -й марки, то индивидуальные значения переводного коэффициента, относящиеся к r -му периоду времени, соответствующему

номеру десятка, равны
$$K_{ir} = \frac{\bar{t}_{ir}}{\bar{t}_{or}}$$

3. Проверяются нуль-гипотеза о независимости выборочных значений коэффициентов от номера печи и нуль-гипотеза о независимости в исследуемом периоде индивидуальных коэффициентов от номера плавки.

4. Если гипотезы подтверждаются, то оценки переводных коэффициентов i -й марки стали отыскиваются по формуле

$$K_i = \frac{1}{n} \sum K_{ir}$$

где n — число коэффициентов в выборке по группе x однотипных печей.

6. Проверяется гипотеза о близости распределения индивидуальных значений коэффициентов к нормальному. Если она под-

тверждается, то среднее квадратическое отклонение \bar{K}_i равно $\frac{\sigma_i}{\sqrt{n}}$

(σ_i — оценка среднего квадратического отклонения K_{ir}).

5.11. По совокупности данных о длительности доводки различных марок стали на семи четырехсоттонных мартеповских печах в качестве основной принята марка Ж, выплавляемая на всех печах в изучаемом периоде. Значения переводных коэффициентов найдены по приведенной выше формуле для изучаемых марок стали и всего периода по S -й ($S=1,7$) печи.

Нулевые гипотезы проверялись методом дисперсионного анализа. Для межгрупповой изменчивости (вызывается различием печей) получены: сумма квадратов $D_1=0,184$; число степеней свободы $K_1=7-1=6$; оценка дисперсии $S_1^2 = D_1/K_1 = 0,0307$; для внутргрупповой (вызывается воздействием всех прочих факто-

ров) — $D_2 = 1,421$; $K_2 = 59 - 7 = 52$; $S_2^2 = D_2/K_2 = 0,0273$; $F = 0,0307/0,0273 = 1,11$.

Табличное значение F_T при $P=0,10$ равно 2,284. Следовательно, нуль-гипотеза не противоречит исходной информации и можно отыскивать общий переводной коэффициент для всех семи исследуемых печей. Нулевая гипотеза о независимости коэффициентов от номера плавки также подтвердилась. Поэтому можно считать совокупность 59 значений индивидуальных коэффициентов по рассматриваемой марке стали случайной выборкой.

Нулевые гипотезы подтвердились для разных марок стали, типов печей и цехов. Получен весьма важный для практики нормирования вывод: хотя абсолютные длительности плавок отдельных марок стали неодинаковы для разных моментов изучаемого периода и разных печей одного типа, их отношения стабильны, устойчивы. Естественно, это справедливо для определенного периода времени, пока резко не изменяются способ производства и основные технические характеристики печей.

Проверка по критерию согласия Пирсона показала, что распределение коэффициентов для каждой марки стали близко к нормальному. Для условий ММК и некоторых ходовых марок стали получены следующие оценки переводных коэффициентов длительности плавок с доверительными интервалами:

Марки	Число плавок	Оценка переводного коэффициента K_i	Ошибка выборочной средней $\frac{\sigma_i}{\sqrt{n}}$	Доверительные интервалы при $P=0,95$
3 сп	162	1,02	0,02	0,99—1,06
4 сп	97	1,03	0,02	0,99—1,07
10 и	59	0,99	0,03	0,93—1,05
45 сп	59	0,92	0,02	0,88—0,98
Ст 1 кп	25	1,11	0,04	1,03—1,19
Ст 2 кп	44	1,09	0,03	1,03—1,15
Ст 3 кп	88	1,01	0,02	0,97—1,05
08 кп	165	1,06	0,01	1,04—1,08
Декапир	41	1,03	0,03	0,97—1,09
10 кп	29	1,01	0,04	0,93—1,10

Доверительные интервалы коэффициентов по многим маркам стали перекрываются. Поэтому целесообразно прибегнуть к вторичной группировке марок по величине коэффициентов. Для рассматриваемого примера получим три группы марок: {45сп}, {3сп, 4сп, 10и, 3ки, декапир, 10кп}, {1кп, 2кп, 08кп} с соответствующими переводными коэффициентами 0,92; 1,02; 1,07.

5.12. Для оценки переводных коэффициентов по редко выплавляемым маркам стали следует использовать интерполяционные формулы, полученные методами множественного регрессионного анализа. В этом случае вместо качественного признака «марка стали» нужно использовать количественные оценки. Если сгруппировать марки стали по набору регламентируемых химических элементов и спо-

собу выплавки, то различие внутри группы будет количественным. Установив для данной группы методами множественной регрессии зависимость длительности плавки от содержания химических элементов в готовой стали, можно затем использовать полученное уравнение как интерполяционную формулу для прямого расчета внутригрупповых переводных коэффициентов. С этой целью одну из марок группы принимают за основную и все коэффициенты уравнения делят на оценку длительности по этой марке.

В качестве примера приведем результаты расчетов по подгруппе хромистых и хромоникелевых сталей (малые печи, свыше 200 плавов). Уравнение регрессии, характеризующее зависимость переводного коэффициента любой марки из группы от содержания (в %) в годной стали элементов C, Mn, Si, P, S, Cr и Ni будет таким: $K = 0,91 - 0,018 C - 0,044 Mn + 0,031 Si + 0,02 P + 3,83 S - 0,013 Cr + 0,085 Ni$; $R = 0,64$.

§ 2. Выбор экономически выгодных режимов управления агрегатом

5.13. Крупный агрегат можно рассматривать как управляемый элемент определенной экономической системы, от состояния которой зависят цели его функционирования. Тогда задача нахождения экономически выгодного режима эксплуатации агрегата сводится к оценке экономической эффективности производственного преобразования ресурсов и к определению способов повышения эффективности в соответствии с заданными целями.

На базе построения экономико-математической модели функционирования агрегата подобная задача решается в три этапа. На первом этапе строится экономическая модель работы агрегата. Изучаются «входы» агрегата — цели, ресурсы, способы управления и «выходы» — показатели и степень эффективности функционирования агрегата. Установление критериев оптимизации завершает построение экономической модели.

Второй этап — математическая формализация задачи. Общая модель крупного технологического агрегата сочетает два типа моделей: экономические — в виде производственной функции и технико-экономические — как зависимости выходных показателей производства от режимных параметров процесса. Третий этап включает экономическую интерпретацию и использование статистических моделей.

5.14. Под производственной функцией понимается соотношение между использованными количествами произ-

водственных факторов (затраты: материальные, на труд, производственные фонды, на совершенствование технологии и организации производства) и количеством конечного продукта при данной технике.

Экономико-математический анализ производственных функций позволяет выявить направления эффективного использования ресурсов. С этой целью рассчитываются следующие характеристики (Хэди, Диллон, 1965):

а) кривые «затраты-выпуск», характеризующие изменение выпуска продукта при меняющейся величине одного из ресурсов, когда затраты остальных закреплены на определенном уровне;

б) предельный дополнительный продукт для каждого ресурса, соответствующий приращению общего продукта от увеличения затрат ресурса на каждую последующую единицу;

в) изокванты, показывающие все возможные сочетания двух ресурсов при данном уровне производства и нормы замещения этих ресурсов.

5.15. Проиллюстрируем расчет экономических оценок с использованием производственной функции на примере моделирования работы крупной электрической печи по производству карбида кальция. В качестве измерителя конечного продукта (функция F) взят показатель выпуска карбида в условных тоннах, т. е. приведенный к стандартному качеству. Для построения производственной функции привлечены факторы, характеризующие материальные затраты: расход кокса (y_1), извести (y_2) и электроэнергия (y_3). Формой функции, наиболее соответствующей логике производственных процессов отдельных агрегатов, признана квадратическая зависимость, включающая парные взаимодействия факторов, т. е. полином вида

$$F = b_0 + \sum_{i=1}^k b_i y_i + \sum_{i=1}^k \sum_{j=1}^k b_{ij} y_i y_j$$

Кривые «затраты-выпуск» рассчитываются для уровней затрат фиксируемых переменных, которые получаются при различных комбинациях следующих значений переменных: \bar{y} , $\bar{y} - \sigma$, $\bar{y} + \sigma$, где \bar{y} — среднее значение, σ — среднее квадратическое отклонение. Сопоставлением всех кривых определяют области, в которых заключены обеспечивающие наиболее высокий выпуск карбида затраты электроэнергии ($\bar{y}_3 + \sigma$), кокса (\bar{y}_1), извести ($\bar{y}_2 - \sigma$).

Таблица 5.1

Пределные дополнительные продукты для отдельных видов ресурсов, т карбида кальция на единицу ресурса

Ресурс	Уровень производства, т карбида кальция		
	73,6	85,1	93,9
Электроэнергия, мВт·ч	0,19	0,32	0,44
Кокс, т	0,24	0,20	-0,05
Известь, т	0,17	-0,01	-0,20

Анализ величин предельных дополнительных продуктов по карбидному производству позволил установить, что с увеличением объема производства (табл.5.1) по мере добавления расхода электроэнергии наблюдается рост отдачи и ее падение — с повышением расхода кокса и извести.

В карбидном производстве из трех ресурсов, включенных в производственную функцию, можно говорить о возможности частичного замещения только двух — электроэнергии и кокса. Уравнение изоквант для электроэнергии, полученное на основе квадратической функции, имеет вид

$$y_3 = 260,34 (0,7607 - 0,0024y_1 \pm \sqrt{0,0076f(y) - 0,8394 \cdot 10^6 y_1 + 0,5437 \cdot 10^6 y_1^2} - 0,55).$$

На основе этого уравнения можно построить сетку изоквант, задавая различные уровни производства, и рассчитать нормы замещения. В табл. 5.2 приведены средние нормы замещения электроэнергии коксом, рассчитанные для двух участков изоквант, соответствующих изменению расхода кокса в интервалах $(\bar{y}_1 - \sigma) \div \bar{y}_1$ и $\bar{y}_1 \div (\bar{y}_1 + \sigma)$.

Таблица 5.2

Средние нормы замещения электроэнергии коксом, мВт·ч/т

Уровень производства, т карбида кальция	Интервал изменения затрат кокса, т	
	36,5—46,5	46,5—56,5
70,8	0,63	0,88
84,2	0,65	0,76
97,5	0,62	0,75

5.16. Анализ производственных функций с помощью перечисленных выше характеристик — предельных добавочных продуктов и норм замещения — позволяет оценить относительный эффект изменения затрат.

На стадии выработки рекомендаций важно не только проследить характер изменения выпуска продукции при разных уровнях затрат того или иного ресурса, но и, сопоставив абсолютный уровень того и другого, сделать выводы о целесообразности их дальнейшего изменения. При этом необходимо каким-то образом соизмерить затраты и результаты. Единственным соизмерителем в этом случае будут цены продукта и ресурса. Сопоставив средние дополнительные продукты с ценами продукта P_F и ресурса P_y , можно установить влияние повышения затрат на прибыль производства.

Таблица 5.3

Средний дополнительный продукт и соотношение цен карбида кальция и ресурсов

Ресурс	Средний дополнительный продукт при интервале усреднения		Соотношение цен
	$\bar{y} - \sigma, \bar{y}$	$\bar{y}, \bar{y} + \sigma$	
Электроэнергия . . .	0,435	0,332	0,103
Кокс	1,141	0,871	0,208
Известь	0,676	0,546	0,158

Повышать уровень затрат на участке между двумя дискретными точками имеет смысл, если $\frac{\Delta F}{\Delta y} > \frac{P_y}{P_F}$, т. е. средний дополнительный продукт превышает соотношение цен. Для рассматриваемого карбидного производства в табл. 5.3 приводятся средние дополнительные продукты, рассчитанные на двух участках, и соотношения цен карбида и ресурсов.

Из сопоставления приведенных данных следует, что увеличение затрат всех ресурсов на обоих участках сопровождается повышением прибыли производства, поскольку во всех случаях средний дополнительный продукт превышает соотношение цен. Но на втором участке, где затраты выше среднего уровня, средний дополнительный продукт по всем ресурсам меньше, чем на первом. Это свидетельствует о тенденции к уменьшению роста прибыли с увеличением затрат.

Изложенный выше прием сопоставления затрат и выпуска может быть использован и для решения вопроса о выгоде замещения одного ресурса другим. В этом

случае нужно сопоставить их среднюю норму замещения в интересующем интервале затрат с соотношением цен

этих взаимозамещаемых ресурсов. При этом если $\frac{\Delta y_i}{\Delta y_j} > \frac{P_{y_j}}{P_{y_i}}$, то замена на рассматриваемом участке ресурса

y_i на y_j выгодна с позиции прибыльности производства.

Так, в карбидном производстве рассчитанное по данным табл. 5.3 соотношение цен кокса и электроэнергии равно 2,021. Из всех средних норм замещения (см. табл. 5.2) максимальная равна 0,88. Поскольку $0,88 < 2,021$, то на любом уровне производства замена электроэнергии коксом при существующих ценах не выгодна, так как приводит к снижению рентабельности производства.

5.17. При анализе и планировании производства могут быть полезны экономические оценки, получаемые на основе производственной функции, построенной по формуле Кобба-Дугласа: $F = \prod_i y_i^{b_i}$. Сумма показателей степе-

ней b_i при переменных в этой функции, получивших название коэффициентов эластичности, выражает сущность зависимости эффективности производства от его масштабов. Если $\sum_i b_i = 1$, то увеличение всех категорий затрат и прирост выпуска (в %) будут одинаковыми. Когда же указанная сумма больше или меньше единицы, выпуск увеличится соответственно в большей или меньшей степени, чем затраты.

В нашем примере по карбидному производству сумма коэффициентов эластичности равна 0,95. Следовательно, увеличение затрат взятых производственных факторов и рост выпуска карбида кальция (в%) должны быть примерно одинаковыми. Этот вывод согласуется с теми оценками, какие получены при анализе квадратической производственной функции, который показал экономическую целесообразность расширения масштаба карбидного производства.

5.18. Экономически выгодное управление агрегатами на базе статистических моделей предполагает конструирование оптимизационной технико-экономической модели, что в первую очередь выдвигает вопрос о выборе критерия оптимальности. Этот выбор обычно осуществляется в следующей последовательности (Черчмен, Акоф, Арноф, 1968): формируется первичный набор целей; устанавливается их иерархия; выбираются технико-экономические показатели (либо другие операционные величины), характеризующие

степень выполнения поставленных целей; решается проблема соизмерения разных целей.

5.19. Набор целей функционирования агрегата формируется под воздействием внешних и внутренних требований к нему. Если агрегат занимает начальное или промежуточное положение в технологической цепочке производства, то к числу факторов формирования целей в первую очередь следует отнести требования последующих технологических переделов, для которых рассматриваемый агрегат производит заготовку или полуфабрикаты. Если работа агрегата есть конечное звено производственного процесса (или, являясь промежуточным звеном, он наряду с полуфабрикатом производит также и готовую продукцию), то важным фактором формирования целей выступают требования, предъявляемые внешними потребителями. И наконец, существенную роль в формировании целей играют и внутренние для системы агрегата экономические требования.

5.20. Установить иерархию целей необходимо для их укрупнения и упрощения. При этом может оказаться, что некоторая цель составляет подцель более общей цели. Например, в первоначальном перечне целей предусматривается снижение себестоимости продукции и в то же время указано снижение удельных расходов сырья и топлива. Понятно, что достижение последних целей есть лишь средство снижения себестоимости и они не имеют самостоятельного значения. Но бывает и так, что снижение удельного расхода какого-либо сырья выступает самостоятельной целью независимо от влияния на уровень затрат, например в случае его дефицитности.

5.21. Для характеристики степени выполнения сформулированных целей могут использоваться различные технико-экономические показатели. Наиболее часто используются показатели производительности агрегата, качества продукции, затрат ресурсов на единицу продукции (общие: себестоимость, расходы по переделу и частные: удельные расходы сырья, топлива, электроэнергии), сумма прибыли, приведенные затраты. Выбор того или иного из них или их комбинации определяется специфическими условиями исследуемого производства.

При выборе из возможного набора критериальных показателей, наиболее ценных для конкретной производственной ситуации, целесообразно определить меру несоот-

ветствия между целью и фактически достигнутыми значениями показателя. Для решения этого вопроса следует прибегнуть к расчету критерия оптимизируемости процесса по каждому из показателей-кандидатов (Кулик, 1964) или использовать экспертные оценки. Для измерения степени достижения качественных целей наиболее целесообразно подобрать количественный эквивалент (например, для характеристики качества продукции использовать показатель химического состава и механических свойств и т. д.). Если это невозможно, используются балльные оценки степени достижения качественной цели либо в качестве измерителя принимается оценка вероятности успешного достижения сформулированной качественной цели. Например, поставлена цель: «выпустить партию изделий без дефектов». Наличие дефектов означает невыполнение цели.

5.22. В зависимости от способа соизмерения целей выделим три варианта многоцелевой оптимизации решения статистической модели.

I вариант. Выбран один основной критерий оптимизации. Другие важные для данных производственных условий технико-экономические показатели работы агрегата выступают в качестве ограничений, по ним задается верхняя или нижняя граница.

II вариант аналогичен первому, за исключением того, что оптимизация осуществляется по обобщенному критерию, который конструируется на основе взвешивания системы частных.

III вариант. Имеется несколько несводимых друг к другу критериев оптимизации одного ранга. Ставится задача найти оптимальный компромиссный режим, обеспечивающий улучшение всех критериальных показателей.

5.23. Оптимизация по условиям I варианта предполагает построение одной модели, состоящей из уравнений регрессии, балансовых и двусторонних ограничений на переменные. Уравнение регрессии, которое описывает влияние независимых переменных на основной критерий, выступает в качестве целевой функции, остальные, отражающие влияние переменных на вспомогательные критерии и взаимосвязь переменных между собой — ограничениями. Оптимизация модели проводится методами линейного программирования с проверкой допустимости полученного оптимального решения (гл. 5). Методика построения статистических моделей рассмотрена в гл. 3.

5.24. Построение оптимизируемой статистической модели по условиям II варианта включает следующие операции:

1) построение для каждого частного критерия уравнения регрессии, описывающего его зависимость от параметров управления технологическим процессом;

2) построение соизмеряющей функции, связывающей частные и обобщенный критерии;

3) построение общей целевой функции;

4) формирование общей системы ограничений на независимые переменные.

Методика построения уравнений регрессии изложена в разделе I. Здесь рассмотрим методику построения соизмеряющих функций, связывающих частные и обобщенный критерии, и конструирование общей целевой функции. Подход к построению соизмеряющих функций зависит от наличия показателя более высокого ранга, с позиций которого выполняется взвешивание, от характера взвешиваемых показателей (количественные и качественные), возможности стоимостной оценки степени достижения целей и др.

Если имеется один основной показатель, не поддающийся непосредственному охвату экономико-статистическим исследованием, и несколько промежуточных более низкого ранга, представляющих его в исходной информации, то соизмерение может быть проведено на основе либо функциональных зависимостей (если таковые имеются), либо регрессионных моделей. При отсутствии основного критериального показателя наиболее целесообразный метод взвешивания — стоимостной. При его использовании предполагается, что можно дать стоимостную оценку всех единиц, в которых измеряются различные цели. Между тем к некоторым целям стоимостная оценка вообще неприменима (например, обеспечение безопасных условий работы), по другим она принципиально возможна, но чрезвычайно затруднена. В этих случаях для соизмерения следует использовать экспертные оценки. Один из методов экспертной оценки относительной значимости множества рассматриваемых целей (весовых коэффициентов K_i) отличается достаточной универсальностью (Черчмен, Акоф, Арноф, 1968).

Например, выбор рационального электрического режима (Р) эксплуатации электросталеплавильных печей проводится по кри-

терию минимума эксплуатационных расходов. Последние по связи с режимом работы печи могут быть подразделены на три основные части. Первая часть — затраты, которые в общей сумме изменяются в основном пропорционально изменению объема производства. Удельная величина затрат не зависит от режима работы печи (H). Вторая часть — относительно постоянные расходы (Y). Удельная величина их изменяется (обратно пропорционально) производительности печи. Третья часть — затраты на электроэнергию (\mathcal{E}), они определяются на 1 т стали и обусловлены режимом работы печи. Такое расчленение эксплуатационных затрат (Z) позволяет ограничиться установленным влиянием параметров режима на

производительность печи ($\frac{q}{t}$, q — вес одной плавки, t — длительность плавки) и удельный расход электроэнергии (W):

$$\frac{q}{t} = f_1(P) \text{ и } W = f_2(P).$$

Для построения связывающей эксплуатационные расходы с производительностью и удельным расходом электроэнергии функции f_3

$$Z = f_3 \left[\frac{q}{t} = f_1(P); W = f_2(P) \right]$$

выразим составные части Y и \mathcal{E} через промежуточные показатели работы печи $\frac{q}{t}$ и W :

$$Y = m \frac{t}{q}; \quad \mathcal{E} = W \cdot C_a,$$

где m — сумма условно-постоянных расходов в расчете на один час работы печи под током; C_a — цена одного киловатт-часа электроэнергии. Тогда

$$Z = H + m \frac{t}{q} + WC_a.$$

Для одной из дуговых электрод печей при выплавке быстрорежущих марок стали получены следующие зависимости показатель работы от величины подводимой мощности:

$$t_{пл} = 302,0 - 10,44 \cdot 10^{-2} P + 12,36 \cdot 10^{-6} P^2;$$

$$W_{пл} = 1243,9 - 4,6 \cdot 10^{-1} P + 6,62 \cdot 10^{-5} P^2.$$

Анализ этих зависимостей показывает, что минимальный удельный расход электроэнергии достигается при значительно меньшей величине подводимой мощности ($P = 3470$ кВт), чем требуется при минимальной длительности плавления ($P = 4230$ кВт). При построении целевой функции приняли, что сумма неизменных затрат на 1 т продукции $H = 15,6$ руб.; сумма условно-постоянных расходов, приходящихся на 1 ч работы печи под током, $m = 26,4$ руб.; вес плавки $q = 10,8$ т, цена 1 кВт·ч электроэнергии $C_a = 0,012$ руб. Подставляя соответствующие значения и заменяя t и W их зависимостями от мощности, получим выражение общей целевой функции:

$$Z = 51,6 - 0,98 \cdot 10^{-2} P + 1,3 \cdot 10^{-6} P^2.$$

5.25. Решение статистической модели в условиях III варианта следует проводить по следующей схеме:

1) построение целевых функций для каждого из выделенных критериев;

2) конструирование общей системы ограничений на независимые переменные;

3) нахождение оптимальных режимов для каждой функции и рационального компромисса между режимами, оптимальными для различных критериев.

5.26. Методику решения задачи рассмотрим на примере многоцелевой оптимизации режима работы карбидной печи одного из заводов синтетического каучука (Проблемы экономико-статистического анализа, 1969).

В качестве критериев оптимизации выступали: максимум выхода полезного продукта из карбида (литражность) — F_1 , максимум производительности — F_2 , минимум себестоимости продукции — F_3 и максимум прибыли — F_4 .

Полезность карбида кальция зависит от содержания в нем чистого продукта, выражаемого показателем качества — литражом (объемом сухого ацетилен в литрах, выделяемого 1 кг карбида кальция). На исследуемом производстве выпускается карбид со средней (литражностью) 260—270 л/кг. Теоретические исследования и уровень мирового стандарта показывают перспективность работы карбидных печей на повышенных литражах (до 310 л/кг). Поэтому целесообразно исследовать процесс на критерий максимума литража.

В связи с тенденцией к химизации народного хозяйства производство искусственного каучука и других органических соединений на основе ацетилена непрерывно растет, что ведет к увеличению спроса на карбид кальция. Поэтому возникает необходимость в критерии максимума производительности печей.

Производство карбида кальция — начальное звено в технологической цепочке производства синтетического каучука. Основные требования к карбидному производству со стороны последующих переделов состоят в выпуске более дешевого карбида при сохранении определенного уровня его качества. Этому требованию удовлетворяет показатель минимума затрат на тонну карбида, приведенного к стандартному качеству.

В связи с тем, что последующий передел — цех ацетилена — имеет ограниченные мощности, при оптимизации

по вышеприведенному критерию необходимо задать ограничение сверху на выпуск карбида. Но эта продукция выступает не только полуфабрикатом, она идет также за пределы завода как товарная. Доля товарного выпуска в общем объеме производства составляет около 60%. Это вынуждает не давать ограничений на выпуск карбида и провести исследование по критерию максимума прибыли.

В качестве целевых функций по F_1 и F_2 можно использовать уравнения регрессии. Для себестоимости и прибыли сделать это не представляется возможным, так как эти показатели учитываются в месячном разрезе, а основные входные факторы — в часовом и сменном. Поэтому модели себестоимости и прибыли конструировались из отдельных элементов. Соизмерение отдельных видов затрат в натуральном выражении осуществлялось на основе действующих цен.

Например, функция сменной прибыли рассчитывалась по такой конструктивной формуле $F_1 = F_2 \cdot P_4 - \sum_{i=1}^5 S_i$,

где S_1, S_2, S_3 — соответственно сменные затраты на электроэнергию, кокс и известь, равные произведению сменных расходов электроэнергии Y_1 , кокса Y_2 и извести Y_3 на их цены (P_i):

$$S_i = Y_i P_i \quad (i=1, 2, 3);$$

S_4 — сменная сумма условно-постоянных расходов, равная $10,38 F_2$, где F_2 — среднее значение сменной производительности печи;

S_5 — пропорциональные расходы по переделу;

P_4 — средняя цена 1 т карбида (с учетом дифференциации продукта по качеству).

Таким образом, построение целевой функции прибыли сводится к получению и последующему соизмерению уравнений регрессии $Y_i = f_i(X)$, где X — множество входных производственных факторов, $i = 1, 2, 3$.

Аналогичным образом строилась функция себестоимости с той лишь разницей, что величины S_i представляли удельные затраты. Кроме того, из рассмотрения были исключены условно-постоянные расходы, приводившие к нелинейности функции, и поэтому рассматривалась функция прямых затрат F_3 .

5.27. Система ограничений для всех критериев включала балансовые, структурные и позитивные ограничения. Первые строились на основе уравнений химических реакций — материального и теплового балансов работы печей

с привлечением в необходимых случаях методов регрессионного анализа. Структурные ограничения, характеризующие внутренние взаимосвязи входных факторов изучаемого процесса, находились как регрессия ведущего фактора в каждой группе от остальных. В число структурных вошли также технологические ограничения на разности между ступенями трансформатора на отдельных фазах. Позиционные ограничения, которые описывают возможный диапазон варьирования факторов, устанавливались на основе технологических норм и имели вид

$$x_{i\min} \leq x_i \leq x_{i\max}.$$

5.28. Для каждой целевой функции методами линейного программирования находятся оптимальные режимы. Задача отыскания компромиссного сочетания режимов, оптимальных по различным критериям, решалась на основе методики, предложенной Х. Юттлером (1967). Она основана на использовании элементов теории статистических решений и линейного программирования. Поиск комбинации наборов режимов эксплуатации, наилучшим образом учитывающей принятые критерии, связан с построением квадратной матрицы потерь. В условиях рассматриваемой задачи она (матрица G^*) имела вид:

x_k	$F_1(x)!$	$F_2(x)!$	$F_3(x)!$	$F_4(x)!$
x_1	0	-0,038	-0,592	-0,31
x_2	-0,07	0	-0,033	-0,024
x_3	-0,0754	-0,001	0	-0,022
x_4	-0,0749	-0,005	-0,098	0

Каждой строке матрицы соответствует один из четырех оптимальных векторов технологических параметров, а каждому столбцу — одна из четырех целевых функций (для $F_i(x) \rightarrow \text{ext}$ принято сокращенное обозначение $F_i(x)!$). Диагональные элементы матрицы — нулевые, остальные меньше нуля или равны ему в случае совпадения режимов для различных целевых функций.

Анализ матрицы G^* показывает, что сравнительно небольшие отклонения от оптимальных значений для всех критериев наблюдаются при выполнении режимов x_2, x_3 и x_4 . Для режима, оптимального с точки зрения литража, выявляются значительные отклонения от оптимальных значений себестоимости (69%) и прибыли (31%). Характерно хорошее совпадение оптимального значения целевой функции $F_2(x)$ (производительности) с ее значениями для

остальных трех режимов (максимальное отклонение 3,8%), что свидетельствует о малой чувствительности этого критерия к разбросу значений параметров технологического режима.

С использованием матрицы потерь решалась задача по определению такой оптимальной стратегии x_0 , при которой потери были бы минимальными по отношению ко всем остальным критериям. Задача оптимизации при этом формулировалась следующим образом. Найти $v = \max!$ при условиях:

$$-0,07 \lambda_2 - 0,0754 \lambda_3 - 0,0749 \lambda_4 \geq v;$$

$$-0,038 \lambda_1 - 0,001 \lambda_3 - 0,005 \lambda_4 \geq v;$$

$$-0,0692 \lambda_1 - 0,003 \lambda_2 - 0,098 \lambda_4 \geq v;$$

$$-0,31 \lambda_1 - 0,024 \lambda_2 - 0,022 \lambda_3 \geq v;$$

$$\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1;$$

$$\lambda_1 \geq 0; \quad \lambda_3 \geq 0;$$

$$\lambda_2 \geq 0; \quad \lambda_4 \geq 0.$$

Получено следующее решение:

$$\lambda_0 = \{0,05; 0,95; 0; 0\}, v = -0,067,$$

т. е. определен оптимально-компромиссный режим $x_0 = 0,05 x_1 + 0,95 x_2$.

Таблица 5.4

Значения целевых функций при различных режимах работы печи

Критерий	Режим		
	действующий	оптимальный для соответствующего критерия	оптимально-компромиссный
Литраж, л/кг	261,2 (237,5—284,9)	307,2 (290,4—324,0)	286,4 (269,6—303,2)
Производительность печи, т	64,1 (34,6—93,6)	87,1 (63,9—110,3)	86,9 (63,7—110,2)
Прямые затраты на 1 т карбида, руб.	51,8 (32,4—71,2)	31,9 (30—51,5)	35,2 (30—54,8)
Сменная прибыль, руб.	1050 (393,6—1706,4)	1663 (1297,4—2028,6)	1593,3 (1227,7—1959,9)

Максимальное отклонение по отношению ко всем критериям оптимальности равно 6,7% при ведении процесса в режиме x_0 .

В табл. 5.4 наряду со средними значениями показателей при различных режимах работы печи дается диапазон возможных колебаний, который рассчитан при вероятности $P = 0,95$ попадания в него оптимальных значений целевых функций.

Выбор оптимально-компромиссного режима сопровождается относительно большими потерями лишь со стороны литража, ухудшение остальных показателей относительно их индивидуальных оптимальных значений незначительно. По сравнению с действующим режимом значения целевых функций в компромиссном режиме улучшаются в среднем: для литража — на 9,1%, производительности — на 33, прямых затрат — на 32, прибыли — на 51%.

Глава 6

МОДЕЛИРОВАНИЕ ОСНОВНЫХ ЭТАПОВ РАЗВИТИЯ ПРОМЫШЛЕННОГО ПРЕДПРИЯТИЯ

6.1. Каждому этапу развития промышленного предприятия (проектирование, строительство, эксплуатация, последний этап, в свою очередь, подразделяется на период освоения, стабильного функционирования и «старения»), затем часто следует реконструкция и т. д.) соответствуют свои модели экономических показателей и задачи. На проектной стадии оценке подлежат проектные характеристики будущего предприятия, в период строительства — показатели этого процесса (удельные капиталовложения, сроки строительства, распределение вложений по срокам строительства), на этапе эксплуатации для каждой стадии строятся специфические модели зависимости экономических показателей от производственных факторов.

Рассмотрим методические подходы к решению задач моделирования производственного процесса на различных этапах развития предприятия (строительство, процесс освоения, режим стационарного функционирования и др.).

§ 1. Оценка уровня и характера распределения капитальных вложений в строительство промышленных объектов

6.2. Для оценки уровня капиталовложений статистическими методами исследовалась зависимость капитальных вложений от характеристик будущего предприятия. При постановке задачи в условиях промышленности строительных материалов (Проблемы..., 1969) исходили из предположения, что затраты на строительство зависят прежде всего от специализации и мощности предприятия. Все остальные факторы (затраты на вертикальную планировку и благоустройство территории, устройство внутриплощадных и внешних сетей, проведение проектно-исследовательских работ и др.) изменяют сметную стоимость строительства предприятия одинаковой мощности в пределах одного экономического района на сравнительно небольшую величину. Подтверждением данного обстоятельства служит факт существования нормативов удельных капитальных вложений, где, по существу, представлена зависимость между мощностью предприятия и затратами на его строительство.

Исходной информацией послужили данные 232 проектных вариантов развития и специализации предприятий сборного железобетона на Урале. Каждый из вариантов характеризовался 18 факторами (общий объем производства и его структура по 17 видам продукции: линейные конструкции, балки подкрановые, стеновые панели и др.), влияющими на общую сумму капитальных вложений и ее элементы по видам производства.

Одно из уравнений, описывающих зависимость общей суммы капиталовложений от исследуемых факторов, имеет вид

$$K = 1353,70 + 54,58x_1 + 4,21x_2 - 22,22x_3 - 8,48x_4 + 13,29x_5 + 22,11x_6 - 13,67x_7 - 17,99x_8 - 17,26x_9 - 35,98x_{10} - 32,42x_{11} - 14,52x_{12} - 12,08x_{13} - 47,78x_{14} + 34,93x_{15} + 8,68x_{16} + 16,06x_{17} - 24,13x_{18}, R^2 = 0,974,$$

где K — общая сумма капиталовложений; $x_1 \dots x_{18}$ — объем производства по отдельным позициям номенклатуры.

(Величина коэффициента детерминации свидетельствует о том, что выбранные факторы объясняют около 97% вариации функции. Подсчет отклонений значений показателя, представленных в исходных данных, от рассчитанных по уравнению регрессии показал (табл. 6.1), что лишь по

пяти наблюдениям отклонения были выше 8%, а среднее отклонение равно 3,2%. В приведенные затраты капитальные вложения входят с коэффициентом $E = 0,17$, поэтому такие отклонения можно считать несущественными.

Отметим, что уравнения были составлены для условий строительства предприятий на территории Уральского экономического района. Формула расчета капитальных вложений при строительстве заводов сборного железобетона в других районах страны будет иметь следующий вид¹:

$$K = A_c / 1,14 (105,28 + 19,12x_2 + 12,51x_3 + 11,50x_4 + 18,88x_5 + 10,51x_6 + 15,13x_7 + 14,53x_8 + 11,82x_9 + 5,81x_{10} + 6,14x_{11} + 16,01x_{12} + 13,93x_{13} + 1,67x_{14} + 43,95x_{15} + 5,61x_{16} + 24,77x_{17} + 16,06x_{18}) + A_{06} (1354,14 + 38,27x_2 + 21,95x_3 + 32,90x_4 + 45,44x_5 + 21,05x_6 + 23,83x_7 + 21,09x_8 + 26,59x_9 + 12,19x_{10} + 18,27x_{11} + 42,45x_{12} + 43,95x_{13} + 37,40x_{14} + 76,39x_{15} + 8,28x_{16} + 59,09x_{17} + 41,78x_{18}),$$

где A_c и A_{06} — территориальные и климатические поправочные коэффициенты на стоимость строительно-монтажных работ и стоимость оборудования соответственно, отражающие удорожание (удешевление) строительства в условиях, отличных от условий центральных районов².

6.3. Определим по этой модели размер капитальных вложений на строительство завода сборного железобетона на территории Волгоградской области при следующей его специализации (тыс. м³):

¹ Для расчетов оказалось удобнее представить общие капиталовложения как сумму затрат на строительно-монтажные работы и оборудование. Коэффициенты детерминации соответствующих уравнений регрессии равны 0,984 и 0,953.

² Нормативы удельных капитальных вложений по промышленности стройматериалов. М., Стройиздат, 1966. 80 с.

Таблица 6.1

Отклонения фактических значений показателя капитальных вложений от рассчитанных по уравнению регрессии

Интервал относительных отклонений	Частота	
	% к итогу	число вариантов предприятий
0,1—1	28,9	67
1—2	23,3	54
2—4	25,0	58
4—6	13,4	31
6—10	7,3	17
10 и выше	2,1	5

Линейные конструкции для промышленных зданий (x_2)	8,16
Балки строительные для промышленных зданий (x_4)	9,02
Стеновые панели для отапливаемых промышленных зданий (x_9)	41,24
Конструкции каркасов для сельскохозяйственных зданий (x_{11})	92,05
Панели покрытий для сельскохозяйственных зданий (x_{13})	29,23

Итого 179,70

Согласно нормативам удельных капиталовложений по промышленности стройматериалов, при строительстве завода сборного железобетона в Волгоградской области $A_c = 1,9$; $A_{об} = 1,0$; размер необходимых капитальных вложений будет равен

$$K = \frac{1,9}{1,14} (105,28 + 19,12 \cdot 8,16 + 11,50 \cdot 9,02 + 11,82 \cdot 41,24 + 6,14 \cdot 92,05 + 13,93 \cdot 29,23) + 1,0 (1354,14 + 38,27 \cdot 8,16 + 32,90 \cdot 9,02 + 26,59 \cdot 41,24 + 18,27 \cdot 92,05 + 43,95 \cdot 29,23) = 7851,01 \text{ тыс. руб.}$$

6.4. Для распределения общей суммы капитальных затрат по годам строительства необходимо установить закономерности распределения объема строительно-монтажных работ во времени и разработать на их основе систему нормативов задела. Решение задачи рассмотрим на материалах строительства объектов химической промышленности (Экономико-статистические исследования, 1969). Исходная информация — помесичное распределение объема строительно-монтажных работ отдельных объектов.

Чтобы обеспечить сопоставимость, объем работ каждого объекта принят за 100%, а общая продолжительность строительства разбита на 16 равных интервалов — исходя из удобства дальнейшего расчета нормативов.

Особенность статистической обработки информации заключалась в необходимости приведения к единому масштабу суммы капиталовложений и сроков строительства по отдельным объектам. Для этого рассчитывались:

1) длина интервала по каждому объекту (x_i) как частное от деления общей продолжительности строительства в месяцах (n_i) на количество интервалов ($s = 16$);

2) величина t как произведение длины интервала по каждому объекту на число интервалов ($k = 1, 2, \dots, 16$), прошедших с начала строительства;

3) накопленный объем строительно-монтажных работ на конец каждого интервала по объектам.

Например, по одному из объектов продолжительность строительства составляет 40 месяцев, освоение средств (в % к итогу) по месяцам строительства: 1—1,2; 2—2,7; 3—1,4; ..., 39—0,3; 40—0,1 — всего 100%. Шаг интервала равен $\frac{40}{16} = 2,5$ месяца, величина t для первого интервала $1 \cdot 2,5 = 2,5$ месяца, для второго $2 \cdot 2,5 = 5$ месяцам и т. д. Накопленный объем строительно-монтажных работ на конец первого интервала составил $1,2 + 2,7 + 0,5 \cdot 1,4 = 4,6$ (%). Итак, в задаче как объем работ, так и время оцениваются в процентах от общего объема и срока строительства.

Логический анализ существа вопроса позволил сформулировать следующие требования к функции:

1. Кривая должна проходить через начало координат, так как при нулевом сроке строительства объем строительно-монтажных работ равен нулю.

2. Первая производная функция неотрицательна: с течением времени объем строительно-монтажных работ либо возрастает, либо остается постоянным, но не может уменьшаться.

3. Кривая должна проходить через точку $t = 16$, $Y = 100\%$: окончанию срока строительства соответствует полное освоение средств.

На основе соображений о характере изменения накопленного объема строительно-монтажных работ во времени и анализа полей корреляции и эмпирических линий регрессии было сделано заключение, что зависимость доли освоения средств на данный момент времени от доли времени, прошедшего с начала строительства, следует искать в виде уравнения кубической параболы. Необходимость учета ограничений обусловила использование методов линейного программирования для нахождения оценок параметров уравнений регрессии.

Полученная в результате решения задачи линейного программирования модель распределения общей суммы капиталовложений во времени имеет вид

$$K_t = 0,12t + 0,912t^2 - 0,033t^3,$$

где K_t — процент освоения средств на момент времени t ; t — число промежутков времени, прошедших с начала строительства.

6.5. Данное уравнение может быть использовано для соответствующих расчетов по строительству объектов химической промышленности. Если, например, проектный срок строительства объекта равен 4 годам (48 месяцам), то освоение средств по годам распределяется следующим образом.

Длительность интервала равна 3 месяцам (48:16). Находим количество интервалов (т. е. значение t), прошедших с начала строительства на конец 1, 2, 3 и 4-го года строительства: 4, 8, 12, 16. Подставляя их в приведенное выше уравнение, получим такое распределение средств по годам (в %):

$$1\text{-й год. } K_4 = 0,124 \cdot 4 + 0,912 \cdot 4^2 - 0,033 \cdot 4^3 = 13,0;$$

$$2\text{-й год. } K_8 = 0,124 \cdot 8 + 0,912 \cdot 8^2 - 0,033 \cdot 8^3 = 42,5;$$

$$3\text{-й год. } K_{12} = 0,124 \cdot 12 + 0,912 \cdot 12^2 - 0,033 \cdot 12^3 = 75,8;$$

$$4\text{-й год. } K_{16} = 0,124 \cdot 16 + 0,912 \cdot 16^2 - 0,033 \cdot 16^3 = 100.$$

Отсюда ожидаемое распределение общей суммы капитальных вложений по годам строительства будет следующим (в %):

1-й год — 13; 2-й год — 29,5 (42,5 — 13); 3-й год — 33,3 (75,8 — 42,5) и 4-й год — 24,2 (100 — 75,8).

Для типичных сроков строительства можно составить таблицы ожидаемого освоения средств, что позволит в большинстве случаев избежать излишних расчетов.

§ 2. Моделирование и прогнозирование процесса освоения производственных мощностей

6.6. Задача сводится к определению наиболее вероятных значений исследуемых показателей в период освоения предприятия на основе некоторых априорно известных характеристик уже построенных и функционирующих предприятий. В ней можно выделить изучение динамики абсолютных значений производственных показателей предприятий и установление соотношения между производственными показателями предприятий различных типов.

Экономико-статистический аппарат находит применение в основном для второго аспекта задачи. Это обусловлено тем, что относительные величины более стабильны во времени и в большей мере зависят от вероятностных характеристик условий работы предприятий, чем абсолютные (на последние значительное влияние оказывают субъек-

тивные факторы — организация работ, уровень квалификации работающих и др.). Для изучения динамики абсолютных величин показателей целесообразнее использовать методы профессионального технико-экономического анализа.

6.7. Общая задача прогнозирования процесса освоения проектной мощности включает ряд типичных частных задач. Основу их классификации составляют два признака: момент прогнозирования и характеристика процесса освоения, подлежащая прогнозу. По первому признаку следует различать прогнозирование процесса освоения объекта до его пуска в эксплуатацию, т. е. на стадии проектирования или строительства, и последующее уточнение прогноза после сдачи объекта в эксплуатацию с учетом имеющейся на момент прогноза дополнительной информации о начале фактического освоения. По второму признаку можно выделить оценку сроков освоения, построение кривых освоения и прогнозирование годового уровня освоения.

6.8. Методику решения задачи прогнозирования кривых освоения рассмотрим на примере угольной промышленности (Вопросы..., 1970). Для анализа использованы два типа фактических характеристик освоения шахт: параметры собственно процесса освоения и производственные характеристики шахт. В качестве первых выступает процент освоения проектной мощности в каждом из первых шести лет с момента пуска шахты в эксплуатацию. Производственные характеристики (32 признака) описывали природно-географические, технико-технологические и организационно-экономические условия работы шахты.

6.9. Экономико-статистическая оценка параметров кривых освоения проводилась по одной из модификаций общей схемы построения дискретно-непрерывных моделей (гл. 3, §4). В ней выделяются следующие этапы.

1. Исследование существующих индивидуальных вариантов кривых освоения, их классификация, формирование обучающей выборки. Метод решения — многомерная классификация шахт по параметрам процесса освоения (выходным характеристикам \vec{Y}). Результат — построение алфавита классов и выбор шахт-представителей каждого класса.

2. Математическое описание полученной на 1-м этапе системы класс-эталонов в пространстве факторов производства (\vec{X}) и обучение распознаванию кривых освоения объектов в нем.

3. Преобразование по результатам обучения пространства факторов производства (\vec{X} в \vec{X}'), которое заключается в минимизации описания класс-эталонов и выборе наиболее эффективной подсистемы признаков (\vec{X}').

4. Свободная классификация (таксономия) шахт в преобразованном пространстве факторов производства (\vec{X}').

5. Построение типичных кривых освоения для сформированных на 4-м этапе классов методами регрессионного анализа.

6.10. На первом этапе задача решалась по индивидуальным графикам освоения. Из всей массы графиков методами распознавания образов выделялись классы наиболее типичных (Елкина, Загоруйко, 1966; Айзерман, Браверман, Розоноэр, 1970). Таких классов при разных способах разбиения получали от 3 до 5. Для дальнейшего анализа была выбрана классификация из 5 классов (табл. 6.2).

Из таблицы видно, что три класса (1, 3, 4) являются основными по численности и имеют наиболее типичные характеристики кривых освоения. К первому классу относятся предприятия с незначительным изменением уровня освоения по годам (15,3—58,9%), к третьему — с равномерным ростом освоения (21,5—107,7%), а четвертый класс характеризуется высоким процентом освоения проектной мощности — от 70% в 1-м году эксплуатации предприятия до 113% в 6-м.

6.11. Полученное на первом этапе распределение исходной совокупности по уровню освоения и представляет собой обучающую выборку для решения задачи второго этапа — формирования описания полученных классов в пространстве производственных факторов (Лбов, 1965).

Таблица 6.2

Характеристика классов

Номер класса	Число предприятий	Средний уровень освоения по годам, %					
		1	2	3	4	5	6
1	33	15,3	40,2	47,2	50,6	54,1	58,9
2	23	36,5	64,9	73,9	76,5	73,7	76,0
3	87	21,5	57,9	78,1	93,3	102,9	107,7
4	28	69,7	90,2	96,8	105,0	107,7	112,7
5	8	69,4	113,6	129,8	133,9	129,5	126,6

Результатом этого этапа будет построение дискретной модели освоения производственной мощности.

6.12. Задача третьего и четвертого этапов — упростить и уточнить модель. Для этого на третьем этапе выделялись наиболее информативные признаки, т. е. существенно определяющие разбиение всей совокупности на классы по характеру кривых освоения. В результате общее число признаков уменьшилось до 8 (т. е. в 4 раза) без снижения точности модели. В их число вошли: годовая проектная мощность, максимальная глубина разработки, угол падения пласта, категория по газу и др. Сравнение классификации по полному набору факторов с классификацией по информативной подсистеме показывает, что при сохранении числа групп в классификациях только пятая часть объектов изменяет свою групповую принадлежность. Такой уровень совпадения классификаций при уменьшении общего набора признаков в 4 раза следует считать удовлетворительным.

6.13. На четвертом этапе выполняется переклассификация шахт в суженном признаковом пространстве с учетом весов признаков и формируется нормативная классификация. Основное требование, предъявляемое к классификации — сосредоточение в выделяемых типичных классах предприятий с наиболее близкими значениями информативных производственных признаков. Для этого опробуются несколько вариантов разбиений, оценивается их качество и выбирается оптимальный вариант исходя из цели исследования и характера информации.

6.14. На пятом этапе для каждого из полученных классов подбирается наиболее подходящая теоретическая кривая освоения³. Анализ результатов расчета показал, что во всех классах кривая освоения наилучшим образом аппроксимируется параболой: по мере увеличения периода с момента пуска предприятия в эксплуатацию темп прироста производственной мощности интенсивно затухает. Одна из моделей имела следующий вид:

$$\text{I класс. } y = -2,9t^2 + 31,5t - 0,3;$$

$$\text{II класс. } y = -3,4t^2 + 36,8t - 0,14;$$

$$\text{III класс. } y = -2,8t^2 + 32,5t + 8,7,$$

³ Расчет проводился по программе Г. П. Ульяновой «Подбор тенденций для динамических рядов», опубликованной в кн.: «Статистические алгоритмы и программы» (Новосибирск, 1970. 290 с.).

где y — процент освоения производственной мощности; t — время с момента пуска шахты в эксплуатацию, лет.

6.15. Модель освоения проектной мощности может быть использована для прогноза применительно к предприятиям, находящимся в периоде освоения. Зная конкретные значения признаков, входящих в информативную подсистему, можно определить принадлежность нового предприятия к одному из трех выделенных класс-вариантов и тем самым его типичную (т. е. наиболее вероятную) кривую освоения.

6.16. После пуска шахты в эксплуатацию можно значительно уточнить первоначальный прогноз. Эта возможность основана на сильной статистической зависимости уровня освоения в данном году от уровня освоения в предыдущий период. Степень тесноты такой связи для одного из классов характеризуется следующей матрицей парных коэффициентов корреляции:

Уровень освоения проектной мощности по годам	y_1	y_2	Парные коэффициенты корреляции			
			y_3	y_4	y_5	y_6
y_1	1,0	0,73	0,64	0,51	0,49	0,45
y_2		1,00	0,86	0,70	0,65	0,62
y_3			1,0	0,87	0,78	0,71
y_4				1,0	0,90	0,83
y_5					1,0	0,94
y_6						1,0

Выявленный характер зависимости позволяет использовать для прогнозирования годового уровня освоения регрессионную модель, в которой аргументами будут не только априорно известные характеристики шахты, но и уровень освоения проектной мощности в предыдущий период. В модель для y_2 помимо априорных характеристик включался y_1 , для y_3 — y_2 , для y_4 — y_1 и y_3 , для y_5 — y_2 и y_4 и для y_6 — y_2 и y_5 . Множественные коэффициенты корреляции для освоения со 2-го года по 6-й соответственно составляют: 0,80; 0,89; 0,89; 0,93; 0,96. Основную информационную нагрузку в этих моделях несут данные об уровне освоения мощности в предшествующие годы. Поэтому краткосрочный прогноз возможен и при использовании чистой авторегрессионной схемы без включения априорных характеристик шахты, что значительно упрощает расчеты.

6.17. Для каждого из сформированных классов в этом случае строится уравнение, выражающее зависимость уро-

вня освоения проектной мощности в t -м году от известного уровня в $(t-1)$ -м:

$$y_2 = f(y_1); y_3 = f(y_2); y_4 = f(y_3); y_5 = f(y_4) \text{ и } y_6 = f(y_5).$$

Параметры регрессии для первых трех уравнений оказались следующими:

	a	b	r_{ij}
I класс шахт			
y_2	34,3	0,83	0,87
y_3	48,8	0,88	0,83
y_4	7,2	1,02	0,90
II класс шахт			
y_2	43,0	0,69	0,67
y_3	25,7	0,84	0,84
y_4	32,8	0,71	0,75
III класс шахт			
y_2	43,2	0,65	0,89
y_3	31,7	0,75	0,84
y_4	22,5	0,86	0,89

6.18. При использовании полученных моделей в прогнозе вначале определяется принадлежность объекта к одному из классов, сформированных по набору производственных факторов, затем рассчитывается ожидаемый уровень освоения по соответствующей регрессии исходя из уровня освоения в предшествующие годы. Основное назначение таких моделей — оперативное уточнение показателей перспективных планов по мере приближения к горизонту планирования. Они могут выступать одним из элементов системы скользящего, непрерывного планирования.

§ 3. Модель функционирования предприятия

6.19. Значение и уровень важнейших экономических показателей работы предприятий определяются не только природными и организационно-техническими условиями, но и положением данного предприятия на траектории его развития. Методику построения модели функционирования предприятия рассмотрим на примере лесозаготовительной промышленности. Функционирование лесозаготовительного предприятия во времени носит отчетливо выраженный по-

этапный характер (Распознавание образов, 1972).

Содержательная интерпретация этапов сводится к следующему. На первом этапе происходит быстрое наращивание производственных мощностей леспромхоза. Предприятие развивается экстенсивно за счет увеличения числа лесопунктов, численности промышленно-производственного персонала, расширения парка машин и механизмов. На втором этапе леспромхоз достигает проектной мощности или приближается к ней. Объем заготовки древесины увеличивается (сравнительно незначительно) в основном за счет роста производительности труда и выявления резервов производства. Верхняя допустимая граница роста производственных мощностей предприятия определяется размером установленной расчетной лесосеки.

На третьем этапе объем заготовки древесины постепенно снижается вследствие ухудшения состояния лесосырьевой базы. Растет износ промышленно-производственных фондов, сокращается численность промышленно-производственного персонала. Активное функционирование предприятия заканчивается, а на его базе очень часто возникает другая организация (лесхоз, межлесхоз и т. п.). Лесозаготовки могут продолжаться, но в сравнительно небольших масштабах.

6.20. Внутри выделенных этапов значения факторов изменяются без нарушения соотношения их значимости в модели, в то время как при переходе предприятия с одного этапа развития на другой структурные соотношения факторов меняются. Это предопределило разбиение всей исследуемой совокупности леспромхозов на три соответствующие этапам группы и описание каждой из этих групп «собственной» статистической моделью, т. е. строилась дискретно-непрерывная модель.

В набор, представляющий исследуемую совокупность леспромхозов Красноярского края, входили следующие характеристики:

1) конечный результат функционирования предприятия (объем вывозки древесины, себестоимость 1 м³ древесины);

2) основные факторы производства (численность промышленно-производственного персонала, величина основных промышленно-производственных фондов);

3) природные условия (средний объем хлыста, среднее расстояние вывозки, породный состав);

4) способ производства (доля вывозки древесины по железной дороге и т. д.);

5) «возраст» предприятия — число лет функционирования — и фактор времени ($t=1,9$).

По каждой из этих характеристик имелись статистические ряды за 9 лет (1962—1970 гг.). В качестве функций рассматривались объем заготавливаемой древесины (y_3), численность промышленно-производственного персонала (y_2), величина основных промышленно-производственных фондов (y_1) и себестоимость 1 м³ древесины (y_4).

6.21. Динамическая модель, описывающая развитие предприятия во времени, представляет собой последовательность соответствующих этапам рекурсивных систем, которые в общем виде записываются так:

$$y_{1t} = b_{10} + \dots + L_{1t} + \eta_{1t}$$

$$y_{2t} = b_{20} + b_{21}y_{1t} + \dots + L_{2t} + \eta_{2t}$$

$$y_{3t} = b_{30} + b_{31}y_{1t} + b_{32}y_{2t} + \dots + L_{3t} + \eta_{3t}$$

⋮

⋮

⋮

$$y_{mt} = b_{m0} + b_{m1}y_{1t} + b_{m2}y_{2t} + \dots + L_{mt} + \eta_{mt}$$

Здесь y_{it} — предсказываемые переменные;

L_{it} — линейные функции предопределенных переменных (взаимонезависимых и входящих с запаздыванием зависимых);

η_{it} — ошибки в уравнениях;

i — индекс предсказываемой переменной;

t — индекс года.

6.22. Для оценки параметров такой системы использовалась программа, реализующая трехшаговый м. н. к.⁴ Кроме оценок коэффициентов в натуральном масштабе она позволяет получить их ошибки σ (среднеквадратичские отклонения), с помощью которых можно еще раз проверить, а в случае необходимости и перестроить структуру моделей. Результаты расчета приводятся в табл. 6.3.

6.23. Заслуживает внимания включение в модель времени как фактора, без чего истолкование результатов

⁴ Рогаткин А. П. Программа трехшагового метода наименьших квадратов. — В кн.: Вопросы построения и применения статистических моделей экономических показателей предприятий. Новосибирск, 1971, с. 213—214.

Оценки систем структурных уравнений

Функция	Свободный член	Основные промышленно-производственные фонды (y_1)	Численность работающих (y_2)	Объем вывозки древесины (y_3)	Объем вывозки древесины в предшествующем году (x_1)	Среднее расстояние вывозки (x_2)
Этап I						
y_1	-103,632 216,438					
y_2	-200,926 184,981					
y_3	137,665 28,086		0,098 0,071		0,418 0,095	-2,769 0,924
y_4	3,586 0,330	0,000279 0,000068				
Этап II						
y_1	215,584 217,880				1,870 0,353	
y_2	151,859 50,321					
y_3	97,956 23,035				0,469 0,070	
y_4	4,878 0,469	0,000175 0,000100				0,0164 0,0036
Этап III						
y_1	114,923 118,976					
y_2	54,010 32,816					
y_3	19,458 31,365		0,125 0,032		0,050 0,019	
y_4	3,196 1,294	0,00115 0,00027		-0,0189 0,0027		0,0300 0,0052

было бы чрезвычайно затруднительным. Каждый коэффициент в уравнении нес бы на себе двойную нагрузку — пространственную и временную, и было бы нелегко определить, вес этих ставляющих, особенно если учесть, что при несовпадении знаков они вообще могут исключать

(натуральный масштаб) и их ошибки

Средний объем хлыста (x_3)	Фактор времени (x_4)	«Возраст» предприятия (x_5)	Характер примыкания (x_6)	Численность работающих в предшествующем году (x_7)	Основные фонды в предшествующем году (x_8)	Множественный коэффициент корреляции (R)
Этап I						
					1,048 0,051	0,951
324,590 182,541		12,785 9,648	2,318 0,648	0,702 0,133	—	0,866
	23,252 5,123		-0,859 0,334			0,916
		0,141 0,053	0,0327 0,0036			0,913
Этап II						
	-47,351 22,199				0,745 0,057	0,847
				0,780 0,044		0,875
			0,592 0,184	0,117 0,019		0,922
-1,613 0,273	0,340 0,031		0,0184 0,0029			0,909
Этап III						
				0,17 0,145	0,555 0,039	0,903
				0,885 0,034		0,926
229,259 42,164						0,686
		0,112 0,033				0,717

друг друга, сводя к нулю роль соответствующего фактора. Подтверждением сказанному могут служить результаты расчетов простейших уравнений себестоимости, построенных с включением фактора времени и без него (табл. 6. 4). Фактор времени впитывает в себя все, что

Таблица 6.4

Оценки параметров уравнений удельной себестоимости

Этап	Модели себестоимости	
	включающие фактор времени	не включающие фактор времени
I	$y_4 = -0,193y_3 + 0,814x_6 + 0,388x_5$	$y_4 = 0,244y_3 + 0,916x_6$
II	$y_4 = -0,256y_3 + 0,513x_2 + 0,404x_5$	$y_4 = -0,338y_3 + 0,576x_2$
III	$y_4 = -0,225y_3 + 0,475x_2 + 0,712x_5$	$y_4 = 0,115y_3 + 0,735x_2$

связано с тенденцией развития лесозаготовительного предприятия.

Включение в исследование «возраста» предприятия дополнительно к фактору времени имеет определенный смысл. Между ними существует тесная взаимосвязь (потому одновременно в одну и ту же модель они не вводились), однако их смысловые нагрузки далеко не одинаковы. Возрастной фактор дифференцирует предприятия с одинаковыми значениями фактора «времени», учитывая тем самым в неявной форме многие качественные моменты, не нашедшие отражения в имеющейся информации.

Таблица 6.5

Оценки параметров при некоторых факторах в модели себестоимости (натуральный масштаб)

Этап	Характер прироста	Объем вывозки древесины	Среднее расстояние вывозки
I	0,0327	—	—
II	0,0184	—	0,0164
III	—	-0,189	0,0300

Необходимость такой дифференциации обусловлена разницей в возрасте предприятий, отнесенных к одному этапу, которая достигает иногда десяти лет.

6.24. Модели этапов развития леспромхозов имеют различную структуру. Значимость факторов от этапа к этапу меняется, и в изменении их влияния на

показатели прослеживаются довольно четкие тенденции. Наглядно видно это из табл. 6.5, где приведены значения некоторых коэффициентов по модели себестоимости, рассчитанные для разных этапов.

Изменение роли различных факторов на разных этапах деятельности леспромхоза вполне объяснимо. Так, от этапа к этапу возрастает значимость фактора «размер предприятия». Преимущество крупных предприятий перед мел-

кими (выражающиеся в строительстве более совершенных дорог, нижних складов, рабочих поселков; в меньшей текучести кадров и т. д.) ощутимы всегда, но особый вес они приобретают на третьем этапе развития предприятия, когда встает вопрос о способах расширения сырьевой базы и о продлении срока действия леспромхоза. Очевидно, что у крупного предприятия материально-технические возможности для решения такого рода задач значительно шире, чем у мелкого.

6.25. Проследим динамику влияния характера примыкания на показатели объема производства и удельной себестоимости. Сплавные предприятия в отличие от прижелезнодорожных характеризуются большим размером заготовок и малым расстоянием вывозки древесины. Фондоёмкость продукции сплавных предприятий за счет невысокой стоимости сооружений ниже, чем прижелезнодорожных. Сплавные предприятия создаются преимущественно в новых районах, поэтому они располагают лучшей сырьевой базой, следовательно, на таких предприятиях более высокая производительность труда, а трудоемкость продукции низкая. Эти факторы непосредственным образом отражаются на удельной себестоимости продукции, которая на сплавных предприятиях в 1,5—2 раза ниже, чем в прижелезнодорожных (имеются в виду I и II этапы развития предприятий).

По мере «старения» лесозаготовительных предприятий, сопровождающегося увеличением среднего расстояния вывозки древесины, различия между этими двумя типами леспромхозов постепенно уменьшаются. На третьем этапе характер примыкания уже перестает играть серьезную роль в формировании показателей работы леспромхоза.

6.26. Полученные результаты можно использовать в различных направлениях. Одно из них связано с возможностью прогнозировать деятельность лесозаготовительного предприятия. Располагая минимумом необходимых сведений (дата освоения леспромхоза, его мощность, остаток эксплуатационного запаса сырья) и установив с их помощью местонахождение данного предприятия на траектории его развития, можно предсказать дальнейший путь его «движения» вплоть до момента закрытия. Более детальный и точный прогноз деятельности леспромхоза осуществляется с помощью полученных «поэтапных» моделей.

ПРИМЕНЕНИЕ ЭКОНОМИКО-СТАТИСТИЧЕСКИХ МОДЕЛЕЙ В МЕЖЗАВОДСКОМ АНАЛИЗЕ И ПЛАНИРОВАНИИ

7.1. На основе статистических моделей предприятий может решаться целый комплекс конкретных задач, связанных с анализом и планированием производства:

- 1) сравнительный экономический анализ работы объектов;
- 2) соизмерение объективных условий производства;
- 3) оценка и прогнозирование экономической эффективности так называемого «горизонтального» технического прогресса, который в конечном итоге складывается из мероприятий, проводимых на отдельных предприятиях, и может быть оценен через модели экономических показателей.

Для решения этих задач может использоваться одна и та же общая экономико-статистическая модель, хотя во многих случаях более целесообразно строить специальные модели. Основное различие состоит в характере использования экономико-статистических моделей.

§ 1. Сравнительный технико-экономический анализ на основе ЭСМ

7.2. Методическую основу технико-экономического анализа составляет сравнительное исследование. Использование ЭСМ позволяет выявить причины отклонения фактического уровня показателей от планового, от уровня прошлых периодов, среднеотраслевого, а также от уровня лучших предприятий. Наибольший интерес представляет сравнение с прошлым периодом, которое в факторном разрезе раскрывает характер изменений в процессе производства и их экономические последствия.

Возможности подобного анализа достаточно широки. С его помощью можно оценить влияние таких факторов, которые не поддаются прямому счету (например, изменение химического состава сырья, использование интенсификаторов и т. п.); «чистое» влияние каждого из этих факторов при закреплении на среднем уровне значений других; просто и быстро получить необходимые конечные

оценки (в частности, при наличии модели и программы ежегодный анализ по министерству потребует нескольких часов работы ЭВМ).

Основная методическая особенность рассматриваемого способа сравнительного анализа — использование пространственной статической модели для анализа динамического процесса. В принципе это обусловлено малым изменением условий производства в короткие периоды времени. Практически адекватность модели обеспечивается, во-первых, ежегодным построением новой модели на основе информации за анализируемый год; во-вторых, использованием разностных форм моделей, когда изменение уровня показателя по сравнению с прошлым годом представляется как функция изменения факторов за тот же период; в-третьих, выбором соответствующего типа модели. Если набор основных факторов и характер их влияния на моделируемые показатели на всех предприятиях одинаков, то регрессионная модель может быть построена по всей совокупности предприятий. В противном случае необходимо выделить типичные группы предприятий и применять дискретно-непрерывную форму модели.

Работы по внедрению системы ежегодного сравнительного анализа на основе статистических моделей разделяются на две части. Первая часть работ выполняется перед внедрением системы в планово-экономическую практику. Она носит поисковый, исследовательский характер — здесь решаются методические вопросы о наборе переменных, способах их измерения и учета, типе модели, границах ее приложимости и т. п. Вторая часть — ежегодный пересчет параметров модели по стандартному набору информации, программам и выбранному типу модели.

7.3. Методику проведения ежегодного сравнительного анализа покажем на примере расчетов по мартовскому производству, предпринятых для оценки влияния изменения производственных факторов на изменение себестоимости мартовской стали в 1967 г. по сравнению с 1966 г.

Расчеты проводились на основе двухклассовой дискретно-непрерывной модели, способ построения которой описан в гл. 3. Данные каждого предприятия по анализируемому и предшествующему годам подставляются в частную модель соответствующего этому предприятию класса. Расчетное значение показателя по данным 1966 г. показывает его возможный уровень в отчетном году при отсут-

ствии изменения условий производства. Разница получаемых оценок показателя характеризует изменение его за счет факторов, включенных в модель. Для каждого предприятия эта разница распределяется по отдельным факторам.

При анализе себестоимости по группе мартеновских цехов, работающих на жидком чугуна, последовательно исследовались, во-первых, характер изменения существенных факторов и их долевое участие в изменении общей суммы затрат и, во-вторых, изменение эффективности производства в отдельных цехах и основные причины этих изменений. Расчеты выполнены в базовых ценах Магнитогорского металлургического комбината и отражают условные изменения затрат под влиянием производственных факторов.

Влияние качества и расхода топлива. В 1967 г. калорийность топлива изменилась (в основном увеличилась) почти по всем цехам. Значительные изменения выявлены в цехах Челябинского завода (на 1,25 и 2,24 ккал), на заводе им. Серова (на 4,08 ккал) и в цехе № 3 завода им. Дзержинского (1,59 ккал). Увеличение калорийности топлива положительно сказалось на величине общих годовых затрат — они уменьшились на 3546,8 тыс. руб.

В свою очередь, с увеличением калорийности топлива удельный расход его соответственно снизился. Уменьшение удельного расхода топлива по 25 цехам (особенно значительное по Кушвинскому (—17 кг/т) и Енакиевскому (—12 кг/т) заводам) снизило себестоимость всего объема производства стали на 10656,81 тыс. руб. В целом по топливу экономия составляет 14203,6 тыс. руб.

Влияние состава сырья. Состав сырья характеризовался долями чугуна и окислителей в шихте. Между обоими факторами имеется зависимость (парный коэффициент корреляции 0,522), поэтому в уравнении регрессии их влияние может взаимно перекрываться и эффект от их действия желательно рассматривать совместно. В одиннадцати цехах доля чугуна в садке в 1967 г. увеличилась, в трех осталась без изменения и по двадцати цехам уменьшилась. Наиболее значительными были изменения на Нижне-Тагильском металлургическом комбинате, в цехах 1 и 2 (—2,3 и —2,5%), на Кушвинском (—3,7%), Коммунарском (—1,4%) и Карагандинском (—2,4%) заводах. По большинству цехов выявилась тенденция к

снижению доли окислителей в шихте. Совокупные изменения по обоим факторам привели к снижению себестоимости на 687,3 тыс. руб.

Влияние годового объема производства. В 1967 г. отмечался рост годовой выплавки стали по большинству предприятий. Особенно значителен он был в цехе 3 Магнитогорского металлургического комбината, на Орско-Халиловском и Череповецком заводах, в цехе 1 Макеевского завода. Общий эффект от роста объема производства по цехам, работающим на жидком чугуна, составляет 5831,41 тыс. руб.

По другим факторам доля в изменении себестоимости рассчитывалась аналогично. По мартеновским печам, работающим на жидком чугуна, получены следующие ее значения (в %):

Факторы увеличения затрат в 1967 г.

Доля стали, разливаемой сверху (x_{18})	27,3
Годовая выплавка спокойной стали (x_{12})	23,9
Простой всего за год на одну печь (x_{21})	16,1
Средний срок действия цеха (x_{22})	10,6
Средний вес плавки (x_7)	9,2
Годовая выплавка углеродистой стали (x_{11})	8,4
Годовая выплавка низколегированной стали (x_{10})	3,9
Содержание куска в ломе (x_{16})	0,6
Итого	100,00

Факторы снижения затрат в 1967 г.

Удельный расход топлива (x_{23})	26,6
Доля чугуна в садке (x_{13})	16,0
Годовая выплавка стали (x_8)	14,6
Количество шихты, заваливаемое одной машиной в год (x_6)	13,1
Калорийность топлива (x_{20})	8,8
Расход кислорода на 1 т стали (x_{24})	8,4
Доля стали, разливаемой в изложницы на тележках (x_{17})	5,6
Содержание в чугуна фосфора (x_{14})	5,1
Доля окислителей в шихте (x_{19})	4,1
Содержание в чугуна кремния (x_{15})	0,1
Итого	100,00

§ 2. Статистическое соизмерение объективных условий производства

7.4. Экономические показатели работы предприятий (\vec{Y}) зависят от большого числа производственных факторов (\vec{X}): объема и структуры основных и оборотных фондов, природных условий, качества сырья, применяемой

техники и технологии, уровня организации труда и производства, квалификации коллектива и т. д. Факторы подразделяются на неуправляемые (\vec{X}), управляемые (\vec{Z}) и случайные (ξ)¹.

К группе неуправляемых относятся факторы, которые не могут быть изменены коллективом предприятия и принимаются им как данные (например, природные условия в отраслях добывающей промышленности; условия, связанные с местонахождением предприятия; установленный объем и ассортимент выпускаемой продукции; условия реализации и др.). Сюда же относятся факторы, которые в некотором периоде с трудом поддаются воздействию: объем и структура основных производственных фондов, структура коллектива предприятия по образованию, возрасту, стажу и другим признакам.

В группу управляемых входят факторы, которые могут быть изменены коллективом предприятия, причем их изменение оказывает существенное влияние на результативные показатели. Сюда относятся уровень организации производства и труда, качество управления, совершенствование технологии, степень использования ресурсов и т. д.

В третью группу объединяются факторы, не поддающиеся измерению, однако оказывающие определенное влияние на результативные показатели производственного процесса: субъективные качества руководителя предприятия, непредвиденные осложнения, метеорологические условия и пр. Влияние этой группы факторов всегда имеет место, но по мере накопления знаний оно уменьшается.

7.5. Возникает задача оценки уровня экономических показателей в зависимости от неуправляемых условий производства при элиминировании влияния управляемых факторов. Для соизмерения объективных условий производства предпочтительнее использовать относительные величины — коэффициенты-соизмерители отдельных показателей эффективности производства. Они позволяют определить, насколько (или во сколько раз) изменится величина экономического показателя при переходе от комплекса неуправляемых условий производства, принятого за базис, к любому другому комплексу производственных условий.

¹ Подобная классификация факторов по степени управляемости дана была еще в работе А. Г. Аганбегяна, В. Ф. Майера «Заработная плата в СССР» (М., 1959).

Обозначив коэффициент соизмерения, соответствующий i -му предприятию (группе предприятий), через K_i , комплекс нерегулируемых факторов предприятия i через \vec{X}_i , аналогичный комплекс для базисных (эталонных) условий через \vec{X}_0 , закрепленный уровень регулируемых факторов через \vec{Z}_i , можно записать

$$K_i^t = [f(\vec{X}_i^t, \vec{Z}_i^t) + \xi] : [f(\vec{X}_0^t, \vec{Z}_i^t) + \varepsilon],$$

где t — период, к которому относится расчет.

Поскольку $f(\vec{X}_i^t, \vec{Z}_i^t) + \xi$ означает $Y_{i\vec{Z}}^t$, а $f(\vec{X}_0^t, \vec{Z}_i^t) + \varepsilon = Y_{0\vec{Z}}^t$, то приведенное выше равенство можно переписать в виде

$$K_i^t = (Y_{i\vec{Z}}^t) / (Y_{0\vec{Z}}^t) + \varepsilon,$$

где $Y_{i\vec{Z}}^t$ и $Y_{0\vec{Z}}^t$ — значения показателей эффективности соответствующего i -го предприятия и базисных условий, зависящие от нерегулируемых факторов, т. е. при закреплении регулируемых на определенном уровне.

При построении K_i для прошедшего периода \vec{Z} закрепляется на среднем уровне, для будущего периода — на проектируемом уровне.

7.6. Соизмерение может проводиться по каждому из нерегулируемых факторов, т. е. строится система частных соизмерителей по факторам. В этом случае коэффициенты-соизмерители представляются системой отклонений и суммируемых поправок (в %) для всей совокупности предприятий:

Δx_i	...	Δx_1	...	Δx_k	Поправка
...
...
$3/\bar{a}_i$...	$3/\bar{a}_1$...	$3/\bar{a}_k$	3
$2/\bar{a}_i$...	$2/\bar{a}_1$...	$2/\bar{a}_k$	2
$1/\bar{a}_i$...	$1/\bar{a}_1$...	$1/\bar{a}_k$	1
0	...	0	...	0	0
$-1/\bar{a}_i$...	$-1/\bar{a}_1$...	$-1/\bar{a}_k$	-1

$-2/\tilde{a}_i$...	$-2/\tilde{a}_j$...	$-2/\tilde{a}_k$...	-2
$-3/\tilde{a}_i$...	$-3/\tilde{a}_j$...	$-3/\tilde{a}_k$...	-3
...
...

Здесь $\tilde{a}_j = (a_j)/\bar{Y}_{oZ}$, \tilde{a}_j — коэффициент линейного уравнения j -го фактора.

Частные соизмерители по факторам сводятся в общий по определенному правилу. Результаты расчетов представляют собой систему суммируемых при линейной форме уравнений (перемножаемых — в случае логарифмически-линейной формы связи) поправок, зависящих от отклонений условий производства от средних. Этим способом можно рассчитать общий соизмеритель для проектируемого предприятия (т. е. для нового комплекса условий).

Возможно и такое соизмерение, когда комплекс нерегулируемых факторов предприятия оценивается одной общей оценкой. В этом случае количество оценок совпадает с числом измеряемых объектов, а предприятия сравниваются в целом по единому показателю.

7.7. Подходы к решению задачи соизмерения объективных условий производства различны. Один из них состоит в том, что соизмерители объективных условий рассчитываются при учете всех факторов по каждому предприятию отдельно. Предприятия, имеющие близкие или одинаковые показатели объективных условий, объединяются в группы, для которых должен устанавливаться одинаковый уровень показателей эффективности. Когда невозможно учесть все факторы, влияющие на показатель эффективности, или учитывается не каждое, а лишь определенное повышение или понижение показателя эффективности, необходим другой подход: группировка однотипных предприятий по определенному набору факторов и соизмерение условий производства по выделенным группам.

Предприятия как в пространстве экономических показателей, так и в пространстве факторов производства распределены неравномерно. В связи с этим используемый метод формирования групп должен обеспечить выделение естественно сложившихся концентраций объектов. Для обоснованного разделения на группы необходимо использовать большое число производственных факторов. Подби-

рать предприятия с совпадающими (или близкими) неуправляемыми факторами производства целесообразно по неуправляемым условиям производства только в определенном смысле, а именно по степени их влияния на величину экономических показателей. Здесь возникает необходимость в предварительном соизмерении влияния неуправляемых факторов на экономический показатель с последующей группировкой по соизмеренным значениям неуправляемых факторов.

При многомерной классификации с предварительным соизмерением признаков группировки число близких предприятий оказывается значительно большим и группы могут быть легко сформированы. Дело в том, что разные наборы неуправляемых факторов производства могут давать одинаковый эффект по влиянию на экономические результаты производства.

7.8. Формирование групп осуществляется двумя способами: группировкой оценок, рассчитанных по регрессионным моделям, и применением многоэтапной схемы распознавания образов. При первом способе строятся уравнение регрессии \vec{Y} по всему комплексу учитываемых факторов (как неуправляемых — \vec{X} , так и управляемых — \vec{Z}). Регулируемые факторы \vec{Z} фиксируются, и рассчитываются «чистые» оценки \vec{Y}_x для каждого предприятия. Далее методами таксономии проводится разбиение совокупности предприятий по полученным значениям \vec{Y}_x . Такой способ удовлетворяет большинству экономических требований, предъявляемых к многомерной классификации предприятий для целей стимулирования. Однако он не лишен ряда недостатков. Основной из них — это ограничение на вид функций, накладываемое методом наименьших квадратов. Предварительный выбор той или иной алгебраической формы связи уже заранее накладывает определенные ограничения, уменьшает свободу классификации. При регрессионном моделировании обычно исходят из предположения о непрерывной форме связи, в действительности же могут иметь место разрывы и скачки. И наконец, методика регрессионного анализа не позволяет прямо использовать качественные признаки.

Второй способ предполагает следующие операции.

1. Совокупность предприятий за каждый год разбивается на одинаковое число групп $A(y)$ по фактическому зна-

чению экономического показателя. Таких разбиений будет несколько.

2. Предприятия, попадающие в одну и ту же группу за все годы или большинство анализируемых лет, принимаются в качестве обучающей выборки, которая охватывает часть первичного массива предприятий.

3. По принадлежности объектов к классам $A(y)$ строится описание этих классов по факторам производства. Получаем классы $A(\vec{X}, \vec{Z})$, и тем самым выполняется неявное соизмерение условий производства.

4. В преобразованном с учетом результатов обучения пространстве факторов производства осуществляется свободная таксономия. Разбиение проводится в пространстве, включающем только нерегулируемые факторы. Математическое описание классов в пространстве нерегулируемых условий производства будет основным для отнесения новых объектов к тому или иному классу.

5. Классификация всех объектов (включая и обучающую выборку) по этому описанию дает решение поставленной задачи — нормативное разбиение $A(x)$, соответствующее искомому $A(y_x)$.

6. По каждой из полученных групп строится внутриклассовая регрессия, которая используется для расчета соизмерителей.

7.9. Расчет соизмерителей объективных условий производства рассмотрим на примере исследований по лесозаготовительному производству. Исходная информация² была представлена данными о деятельности 74 лесозаготовительных предприятий Урала и Сибири за 8 лет. Производственные факторы подразделялись на ряд групп:

а) местоположение предприятия и качество используемых лесосырьевых ресурсов;

б) объем, структура и качество производственных фондов;

в) транспортные условия (удаленность и характер примыкания нижних складов, вид используемого транспорта);

г) централизованно регламентируемые экономические условия деятельности предприятия (объем и структура производства и т. д.).

² Исходная информация была представлена Тюменским научно-исследовательским институтом лесной и деревообрабатывающей промышленности (НИИПлесдрев).

Факторы первой из перечисленных групп — нерегулируемые по своей природе, остальные следует отнести к неуправляемым на уровне предприятия. Изменение производственных фондов, характера транспортных связей и т. п. требует значительных затрат средств и времени, и для отдельного периода (года) их можно рассматривать как заданные, сложившиеся. Задания по объему и номенклатуре продукции, действующие отпускные цены, надбавки к заработной плате и т. п. — нерегулируемые факторы в пределах срока действия соответствующих юридических положений.

В качестве показателей экономической эффективности рассматривались производительность труда (комплексная выработка) и себестоимость.

Предприятия группировались вторым способом по значению экономических показателей на основе информации за 8 лет. Метод таксономии (Елкина, Загоруйко, 1966; Елкина, 1971) применялся к одномерной совокупности объектов, где каждый объект был представлен несколькими годовыми реализациями (в нашем случае 592 реализации). По уровню выходных показателей (выработка — в м³, себестоимость — в коп.) сформированные группы предприятий можно охарактеризовать следующим образом:

1) предприятия высшего уровня, имеющие наиболее высокий показатель производительности труда (539—944) и низкий — затрат на производство (382—540) — 39 годовых реализаций.

2) предприятия низкого уровня — с малым значением производительности труда (259—513) и высоким — себестоимости (875—1113) — 113 годовых реализаций;

3) предприятия среднего уровня, имеющие средний показатель себестоимости заготовки древесины (542—851) и высокий — производительности труда (534—944) — 100 годовых реализаций;

4) предприятия со средним уровнем себестоимости (542—851) и низким — производительности (259—513) — 340 годовых реализаций.

По числу годовых реализаций наиболее представительны вторая, третья и четвертая группы. Предприятия по группам распределялись так: в первую группу попало 2 предприятия, во вторую — 14, в третью — 48 и в четвертую — 10. Эти предприятия типичны для группы в том смысле, что большинство их годовых реализаций относятся к данной группе.

Формирование обучающей выборки предполагает проверку устойчивости (представительности) полученных групп. Под временной устойчивостью группы понимается сравнительная стабильность ее состава в анализируемый период времени, т. е. из года в год набор предприятий в группе мало изменяется. Уровень устойчивости предприятий может оцениваться величиной μ_j .

Если число годовых реализаций обозначить через S , то каждому предприятию j можно поставить в соответствие S -мерный вектор

$$K_i^j = (K_1^j, K_2^j, \dots, K_S^j), j = 1, \bar{N}.$$

Величина $K_i^j (i = 1, \bar{S})$ обозначает, что данное предприятие попадает в один и тот же таксон (группу) K_i^j лет. Понятно, что

$$\sum_{i=1}^S K_i^j = S.$$

Мера неопределенности (H_j), возникающая при отнесении предприятия к классу устойчивых, может быть рассчитана по выражению

$$H_j = - \sum_{i=1}^S K_i^j / S \log_S K_i^j / S.$$

В таком случае показателем устойчивости предприятия во времени служит величина $\mu_j = 1 - H_j$.

Обоснованием для такого выбора меры μ_j служит тот факт, что $0 \leq \mu_j \leq 1$ принимает значение 1, когда все годовые реализации попадают в один таксон, и значение 0, если все годовые реализации попадают в разные таксоны.

Распределение предприятий по их временной устойчивости показано в табл. 7.1. К устойчивым представителям групп следует отнести предприятия с мерой устойчивости μ_j не ниже 0,57. Всего к категории устойчивых отнесено 61 предприятие, или 82% от общей совокупности исследуемых объектов. Остальные 13 предприятий относятся к типу неустойчивых и характеризуются ситуациями 5, 1, 1, 1, 1; 4, 3, 1, 0; 4, 2, 2, 0 и 3, 3, 2, 0. Для них характерно распределение годовых реализаций по разным группам с малой величиной вероятности.

Анализ состава групп позволил установить, что дальнейшее исследование необходимо проводить по трем группам:

имеющей самый высокий уровень комплексной выработки (534—944 м³) и средний уровень себестоимости (542—851 коп.), типичные предприятия которой по уровню работы являются лучшими в рассматриваемой совокупности;

имеющей средние выходные показатели в данной совокупности предприятий (наиболее многочисленная и типичная);

имеющей выходные показатели ниже среднего уровня — высокую себестоимость заготовки древесины (от 875 до 1113 коп.) и низкую комплексную выработку (259—513 м³).

Первую группу необходимо исключить из дальнейшего исследования ввиду нетипичности ее представителей.

Таблица 7.1
Распределение предприятий по группам в пространстве выходных показателей с учетом устойчивости годовых реализаций

Ситуация	Всего предприятий по группам					1 группа	2 группа	3 группа	4 группа	Всего предприятий	
	1	2	3	4	5						
Мера устойчивости	1,00	0,82	0,73	0,64	0,68	0,57	0,48	0,67	0,53	0,49	0,48
1 группа		2(III) 3(IV)	1(III) 1(I, III)	1(I, III)	1(I, III)	1(I, III)	1(II, III, IV)	1(III)	3(I, III) 1(IV, III)		2
2 группа					6(IV)	2(II)					14
3 группа	40	3(0) 4(IV) 1(II)	2(0) 2(II) 8(IV)	1(I, II, IV)	2(II, I) 1(IV, I) 1(IV, II)	1(I, I)	1(I, II)	1(I, II, IV)	1(I, II) 1(II, IV) 1(IV, I) 2(II, I)	1(I, II) 1(IV, II)	48
4 группа		1(II)	1(II, III)	1(II, III)	5(III)	1(III, I)			2(III)		40
Всего предприятий	40	14	43	2	13	6	2	3	9	1	74

Примечание. В скобках указаны номера групп, к которым относится меньшинство годовых реализаций. Например, 1(II, III) означает, что 5 годовых реализаций одного предприятия попали в первую группу, 2—во вторую и 1—в третью.

Матрица переходов годовых реализаций предприятий при переклассификации в пространстве производственных факторов

Группа в пространстве \vec{Y}	Группа в пространстве \vec{X}			Всего переходов
	A	B	C	
1	1 реализация 4	1 реализация 10 2 реализации 2 3 реализации 3 4 реализации 1	5 реализаций 1 2 реализации 1 1 реализация 1	39
2		1 реализация 5	1 реализация 3 2 реализации 2	12
3	1 реализация 1			1
4	1 реализация 7 2 реализации 5	1 реализация 11 2 реализации 7 3 реализации 1		45
Всего переходов	22	60	15	97

Примечание. «1 реализация 4» означает, что из группы 1 в группу A перешла 1 годовая реализация по четырем предприятиям.

ного его объективными факторами. Переход же из 4-й группы в A и B характеризует работу предприятий ниже уровня, определяемого объективными условиями их деятельности.

Анализ переходов предприятий подтверждает заключение о нетипичности для рассматриваемой совокупности предприятий 1-й группы в пространстве выходных показателей, характеризующейся высоким уровнем работы. Перегруппировка предприятий в пространстве факторов производства позволила установить, что подавляющее большинство объектов исследуемой совокупности работало на уровне, обусловленном объективными возможностями их

После выделения групп-эталонов по значению результативного показателя (y) осуществляется переход к описанию этих же групп в пространстве факторов производства (\vec{X}). Ведь результативный показатель используется только для получения первичных групп предприятий. Чтобы получить окончательные группы, необходимо учесть влияние неуправляемых факторов производства. С переходом в пространство \vec{X} для соответствующих класс-эталонов необходимо задать описание — некоторую функцию от элементов, входящих в класс и необходимых для формирования окончательных групп. Эта функция есть мера близости, показывающая близость любого произвольного элемента с данным класс-эталонном. При обучении и на следующих этапах работы применялись алгоритм и программа, разработанные Г. С. Лбовым (1965).

По полученному в процессе обучения описанию классов в пространстве \vec{X} осуществляется переклассификация всех предприятий. В результате для некоторых объектов может измениться номер группы. Если при исследовании был учтен достаточно полный набор неуправляемых факторов, то переход из группы в группу можно трактовать как отклонение качества работы коллектива предприятия от среднего уровня.

Здесь возможны три варианта. Первый: предприятие не изменяет своего относительного положения (номера групп в пространстве Y и \vec{X} совпадают). Это означает, что качество работы коллектива предприятия находится на среднем уровне. Второй: номер группы у предприятия в пространстве Y больше, чем в пространстве \vec{X} . Можно сделать вывод, что коллектив предприятия работает ниже своих возможностей. Третий вариант противоположен второму. При этом степень отклонения от средних условий работы пропорциональна числу перешагиваемых групп. Матрица переходов годовых реализаций предприятий приведена в табл. 7.2.

Общее число годовых реализаций предприятий, перешедших из одних групп в другие, равно 97. Из них 39 годовых реализаций перешло из 1-й группы, представленной нетипичными предприятиями, преимущественно в группы B и C. Остальные переходы годовых реализаций были из 2-й группы в группу C и из 4-й группы в группы A и B. Переход из 2-й группы в группу C означает, что предприятия в отдельные годы работали выше уровня, обусловлен-

Таблица 7.3

Фактор	Вероятность	Вес*
Средний запас на 1 га, м ³	0,218	0,345
Удельный вес примыкания складов к железной дороге, %	0,168	0,266
Доля лиственницы, %	0,148	0,234
Доля ели-пихты, %	0,128	0,203
Доля березы-осины, %	0,108	0,171
Удельный вес лесозаготовок, %	0,028	0,046

* Вес фактора — условная вероятность его выбора в предположении, что фактор входит в эффективную подсистему признаков.

работы. В отдельные годы и по отдельным предприятиям имели место отклонения результатов работы от обуславливающих их факторов производства.

Использовать для описания групп полный набор неуправляемых факторов в ряде случаев оказывается практически неудобно, поэтому возникает необходимость в минимизации описания — выборе эффективной подсистемы признаков. Алгоритм случайного поиска с адаптацией (Лбов, 1965) в качестве критерия эффективной подсистемы признаков рассматривает минимальное число ошибок отнесения в классы и минимальное взвешенное число ошибок, если задана матрица потерь. Результаты выбора эффективной подсистемы признаков по этому алгоритму приведены в табл. 7.3.

Наибольшую вероятность имеет признак «средний запас на 1 га», затем следуют «удельный вес примыкания складов к железной дороге» и «состав лесонасаждений». Малой вероятностью отличается признак «удельный вес лесозаготовок». Факторы, не вошедшие в эффективную подсистему, имеют вероятности, равные нулю или близкие к нему.

Чтобы для каждой из полученных групп предприятий найти зависимости выходных показателей от факторов производства — параметры непрерывных внутригрупповых моделей, можно использовать два подхода. Первый заключается в построении непрерывных моделей для каждой группы без учета вероятности принадлежности отдельных предприятий к той или иной группе. При втором подходе вероятности учитываются и модель, например, для себе-

стоимости будет следующая:

$$A \quad y = 868,1 + 2,185x_1 + 0,429x_2 - 170,2x_6 + 1,179x_8 - (542 - 851 \quad -0,221x_9 + 0,266x_{13};$$

534—944)

$$B \quad y = 911,0 - 0,116x_2 - 21,49x_6 + 1,823x_8 - 0,215x_9 - (542 - 851 \quad 0,899x_{10};$$

259—513)

$$C \quad y = 1032,2 - 0,529x_1 + 2,424x_3 + 0,141x_5 - 379,3x_6 - (875 - 1113 \quad -1,474x_8 - 0,116x_9 + 21,47x_{14},$$

259—513)

где x_1 — глубина снежного покрова;

x_2 — удельный вес сосны;

x_3 — удельный вес ели с пихтой;

x_4 — удельный вес лиственницы;

x_5 — удельный вес березы с осиною;

x_6 — средний объем хлыста;

x_8 — среднее расстояние вывозки;

x_9 — объем вывозки;

x_{10} — удельный вес примыкания к сплаву;

x_{13} — удельный вес вывозки автомобилем;

x_{14} — удельный вес вывозки тракторами.

Частные соизмерители по факторам образуют систему отклонений и суммируемых поправок для всей совокупности предприятий. Приведем такую систему, рассчитанную по одному из уравнений для себестоимости:

Глубина снежного покрова	14,6	7,3	0	-7,3	-14,6
Удельный вес насаждений:					
сосны	-20,8	-10,4	0	10,4	20,8
лиственницы	-23,6	-11,8	0	11,8	23,6
березы-осины	47,0	23,5	0	-23,5	-47,0
Примыкание складов:					
к сплаву	54,0	27,0	0	-27,0	-54,0
к пункту потребления	21,0	10,5	0	-10,5	-21,0
Объем вывозки	-135,4	-67,7	0	67,7	135,4
Районный коэффициент	0,08	0,04	0	-0,04	-0,08
Поправка, %	4	2	0	-2	-4

Эта система позволяет оценить степень отклонения себестоимости от ее среднего значения в зависимости от аналогичных отклонений любого из перечисленных факторов. Зная условия производства данного предприятия (на-

Таблица 7.4

Фактор	Средние условия производства	Условия производства данного ЛПХ	Отклонение от средних условий	Поправка, %
Глубина снежного покрова	57,7	104,0	46,3	12,5
Удельный вес насаждений:				
сосны	31,0	63,0	32,0	-6,0
лиственницы	2	0,1	-1,9	0
березы-осины	32,4	15,4	-17,0	-1,5
Примыкание складов:				
к сплаву	9,4	87,9	78,5	6,0
к пункту потребления	39,1	12,0	-27,1	-5,0
Объем вывозки	316,4	318,7	2,3	0
Районный коэффициент	1,15	1,7	0,55	28,0
Значение коэффициента-соизмерителя				34,0

пример, Ханты-Мансийского леспромхоза) и средние условия, находим отклонения от последних (табл. 7.4). По значениям отклонений соответствующих условий определяется величина поправок по таблице 7.4 для каждого фактора. Так, отклонению по фактору «глубина снежного покрова» (46,3) соответствует величина поправки 12,5%; по фактору «удельный вес сосны» (32,0) — (-6%) и т. д. Суммируя все поправки, получаем общий коэффициент для леспромхоза, равный 34%.

Соизмерители, оценивающие объективные возможности каждого конкретного предприятия, могут быть получены как непосредственно на основе моделей, так и с помощью системы отклонений и суммируемых поправок. В качестве иллюстрации приведем результаты расчетов коэффициентов-соизмерителей предприятий по модели себестоимости за 3 года (табл. 7.5).

Из таблицы видно, что коэффициенты-соизмерители по себестоимости сравнительно стабильны во времени (для большинства предприятий изменение не превышает 5%), величина их в выборке колеблется в пределах 0,93—1,28. Общий диапазон их изменения по предприятиям, входящим в анализируемую систему (Урал и Западная Сибирь), составляет 0,73—1,37. Таким образом, перегулируемые условия производства на предприятиях значительно различаются.

Таблица 7.5

Объединение, леспромхоз	Коэффициент по себестоимости*			
	1-й год	2-й год	3-й год	Среднее значение
Серовлес				
Карпинский	0,98	0,94	0,99	0,97
Тагиллес				
Нижне-Лялинский	1,21	1,32	1,30	1,28
Свердлес				
Вогульский	0,98	1,03	0,99	1,00
Полевской	0,93	0,98	0,95	0,95
Талицкий	1,01	0,93	0,94	0,96
Шамарский	0,99	1,06	1,02	1,02
Алацаевсклес				
Гаранинский	0,99	1,00	0,99	0,99
Ясашинский	1,15	1,22	1,18	1,18
Тавдалес				
Карабашинский	1,06	1,09	1,09	1,06
Тюменская обл.				
Землеборский	0,94	0,93	0,90	0,93
Кондинский	1,19	1,14	1,03	1,11
Омская обл.				
Усть-Ишимский	1,01	1,03	1,02	1,02
Томская обл.				
Берегаевский	1,03	0,97	0,98	0,99

* Коэффициенты по себестоимости представляют собой отношение базисного показателя к показателю *i*-го предприятия.

Коэффициенты-соизмерители могут использоваться для сравнительного анализа работы предприятий, объединений и т. д., а также в плановой практике при установлении правильных соотношений по объему производства, производительности труда, затратам на производстве и т. п.

§ 3. Статистическая оценка технического уровня производства

7.10. Технический уровень производства — общий результат непосредственных проявлений технического прогресса на некоторый фиксированный момент времени для определенного объекта.

Существует несколько точек зрения по поводу того, комплексом каких показателей измерять технический уровень производства. Так, С. М. Ямпольский и В. Г. Чирков (1971) предлагают использовать такие показатели, как удельный вес активной части основных фондов, средний возраст оборудования, удельный вес морально устаревшего оборудования, фондовооруженность труда, средний «возраст» технологических процессов, потенциальная электровооруженность; коэффициент технического развития, выражающий уровень механизации и автоматизации производства; степень отдачи трудовых и материальных ресурсов предприятия, уровень качества выпускаемой продукции. В работе Б. Ф. Зайцева и В. Г. Чиркова (1972) технический уровень производства описывается показателями технической характеристики средств труда, вооруженности труда, уровня технологии, уровня используемых предметов труда, механизации и автоматизаций производства.

О. И. Волков (1970) относит к техническому уровню производства характеристики уровня механизации и автоматизации производственных процессов, электровооруженности. По нашему мнению, комплекс характеристик технического уровня должны составлять те показатели, изменения которых вызваны реализациями технического прогресса на промышленном предприятии.

В общем случае к ним следует отнести:

- 1) размер, структуру и качество основных фондов;
- 2) вооруженность труда;
- 3) уровень механизации и автоматизации производства;
- 4) качество материальных оборотных фондов;
- 5) уровень используемых производственных способов;
- 6) качество выпускаемой продукции.

Конкретизация этого комплекса показателей должна

осуществляться для каждой отрасли и каждого технологического способа отдельно с учетом их специфики.

7.11. Выделим три особенности показателя технического уровня производства:

комплексность и многомерность;
неоднородность признаков (как количественных, так и качественных), входящих в систему характеристик технического уровня;

разное содержание, информационную ценность и направления развития отдельных характеристик.

Соизмерять комплекс разнородных характеристик можно либо с помощью обобщенных оценок самого технического уровня объекта, либо через его влияние на экономические показатели производства.

7.12. Первый подход предполагает использование методов многомерной классификации. При статической постановке задачи (относительное измерение технического уровня на предприятиях в определенный момент времени) осуществляется многомерная группировка объектов совокупности по комплексу характеристик, описывающих технический уровень. Можно утверждать, что в результате такой группировки классы будут получены, поскольку характеристики технического уровня взаимосвязаны и выражают общность технико-технологических элементов производственного процесса. Классы формируются из предприятий, близких по значениям характеристик технического уровня. Полученные классы ранжируются по техническому уровню методом экспертных оценок. Здесь возможны два способа. Один из них заключается в том, что эксперты ранжируют характеристики технического уровня по каждому классу. На основе индивидуальных экспертных оценок определяется групповая оценка технического уровня класса, по которой затем проводится ранжировка классов. Другой способ состоит в определении «положительных» направлений изменения признаков технического уровня: по каждому признаку определяется то направление изменения, которое соответствует его совершенствованию под влиянием технического прогресса. Тем самым пространство признаков определенным образом упорядочивается — вводятся направленные оси. Затем в каждом классе находится среднее значение каждого признака, классу ставится в соответствие набор средних значений признаков, выделенных в систему характеристик техни-

ческого уровня. Сравниваются все пары классов по этому набору, а от попарного сравнения переходят к полному упорядочению классов.

В динамической постановке задача соизмерения разнородных характеристик сводится к анализу во времени элементов статистической классификации. Исследуется изменение классификаций предприятий по техническому уровню за несколько лет: изменение численности классов за счет перемещения объектов из класса в класс и появления новых объектов, исчезновения старых, анализ движения характеристик существующих классов. В отношении каждого объекта совокупности определяется характер его движения. Здесь различают движение объекта вместе с классом, что соответствует плавным изменениям технического уровня, эволюционной форме технического прогресса, и переходы объекта из класса в класс, что означает качественное изменение технического уровня, скачкообразный тип технического прогресса:

7.13. Задача экономической оценки технического уровня с точки зрения его влияния на результативные экономические показатели предприятий (второй подход) аналогична соизмерению объективных условий производства. Отметим некоторые особенности. При выборе экономических показателей, выступающих в качестве критерия соизмерения комплекса характеристик технического уровня, необходимо иметь в виду, что экономический эффект технического прогресса в отрасли распадается на две составляющие: эффект у производителя от применения прогрессивной технологии, орудий и предметов труда и эффект у потребителя, получающего более прогрессивную, лучшего качества и более дешевую продукцию. Отражение той или иной составляющей при выборе показателей экономической эффективности зависит от двух обстоятельств.

Во-первых, от положения, которое занимает объект исследования в технологической цепочке общественного производства, т. е. от того, к добывающей или обрабатывающей отрасли относится предприятие. На предприятии добывающей отрасли технический прогресс проявляется не столько в качестве производимой продукции (здесь очень многое определяется природными условиями), сколько в совершенствовании технологического способа добычи и условий его реализации. Для предприятия обрабатывающей промышленности более важным направлени-

ем технического прогресса является повышение качества выпускаемой продукции.

Во-вторых, выбор показателей, используемых для экономической оценки технического уровня, зависит от целей его измерения. Если, например, такая оценка проводится для решения задач отраслевого планирования, то рассматривается только процесс производства. Когда технический уровень оценивается для целей программного планирования и управления комплексом крупных работ, необходимо рассматривать обе составляющие эффекта, так как от технического уровня производимой на одном этапе продукции зависит производство на другом этапе.

В настоящей работе проблема оценки влияния технического уровня на экономические показатели рассматривается в рамках отраслевого планирования. Наибольшее внимание уделяется эффекту, реализуемому производителем. Эффект у потребителя учитывается путем оценки влияния потребительских свойств продукции на издержки производства. Наличие разнородных признаков, требование учета независимости признаков и однородности совокупности определяют решение задачи методами дискретно-непрерывного моделирования.

7.14. Соизмерение технического уровня через его влияние на экономические показатели рассмотрим на примере угольных шахт Кузнецкого бассейна (Распознавание образов, 1974). В качестве показателей экономической эффективности были выбраны производительность труда и себестоимость добычи угля. В силу специфики угольной промышленности рассматривался комплекс характеристик горно-геологических условий и их связь с элементами технико-технологической структуры.

Наиболее важными горно-геологическими факторами, определяющими выбор технологической схемы (главным образом системы разработки) и применяемой техники, являются мощность и угол залегания пластов угля. Кроме того, на технико-технологическую структуру шахты и уровень экономических показателей оказывает влияние еще ряд горно-геологических факторов: газоносность, водообильность, устойчивость боковых пород, глубина разработки, количество пластов, нарушенность пластов и т. д. Факторы (например, устойчивость пород, нарушенность пластов), измерители по которым отсутствовали, в исследовании не включались.

Система характеристик и измерителей технического уровня для угольной шахты формировалась исходя из общего представления о техническом уровне производства на промышленном предприятии (Волков, 1970; Зайцев, Чирков, 1972; Ямпольский, Чирков, 1971) и специфики производственного процесса на угольной шахте (Будницкий, 1959).

В качестве измерителей технического уровня были приняты следующие показатели работы шахты: характеристики вооруженности труда, технологии и уровня механизации; технические показатели, характеризующие производственный процесс в целом, с упором на очистный забой. Общий первоначальный набор признаков включал 36 характеристик исследуемого процесса. Из них признаки x_1-x_{13} описывают горно-геологические условия производства, $x_{14}-x_{28}$ — характеристики технического уровня, $x_{29}-x_{33}$ — организационно-экономические условия производства.

На первом этапе исходная совокупность шахт разбивалась на два класса по применяемой технологии добычи угля. Система разработки — один из основных признаков технологического способа подземной добычи угля. То есть указанная классификация совокупности предприятий выполнялась по одной из самых важных характеристик технического уровня производства. С другой стороны, такое разбиение учитывает и горно-геологические условия, поскольку определенная система разработки применяется в соответствующих природных условиях.

На втором этапе в каждой из полученных подсовокупностей проводилась многомерная классификация по комплексу остальных характеристик технического уровня. На третьем этапе внутри каждого класса строилось уравнение зависимости экономического показателя от факторов производства. Закрепляя прочие факторы, кроме характеристик технического уровня, на стабильном для класса уровне, получали условную экономическую оценку технического уровня шахты. В качестве эталона принималась «средняя» шахта, и рассчитывались оценки в виде относительных величин. В табл. 7.6 приведены оценки технического уровня по показателю производительности труда для одного из классов.

Имея оценку технического уровня для группы предприятий с одним технологическим способом производства,

сравнительно давно освоенным, можно оценить экономическую эффективность нового технологического способа, представленного единичными промышленными объектами, сравнительно недавно введенными в действие. Таким, например, является гидравлический способ добычи угля, представленный в настоящее время четырьмя шахтами Кузнецкого бассейна.

Для экономической оценки гидродобычи была сконструирована модель экономического показателя на всей совокупности «сухих» шахт. Подставляя в нее значение характеристик технического уровня лучшей «сухой» шахты и значения характеристик природных условий и организационно-экономических факторов гидрошахты, получили результат, возможный на лучшей «сухой» шахте в условиях, в которых реально находится гидрошахта. Сравнение его с фактическим показателем гидрошахты позволило судить об эффективности гидравлического способа добычи угля. Расчеты показали, что гидрошахта Полысаевская-Северная, разрабатываемая как «сухая» шахта, имела бы показатель производительности труда 84,9 т/чел, тогда как фактическая производительность на этой шахте составляет 103,4 т/чел.

§ 4. Пример использования ЭСМ в оптимальном отраслевом планировании

7.15. Лесозаготовительная промышленность Красноярского края была постоянным объектом моделирования на протяжении последних лет. И это не случайно. Лесозаготовка и деревопереработка здесь очень развиты и имеют самые благоприятные перспективы. В ближайшем будущем вывоз древесины за пределы края будет сведен к минимуму. Возрастет комплексность и глубина переработки древесины. Все это полностью оправдывает выделение

Таблица 7.6
Экономическая оценка
технического уровня шахт
по показателю производи-
тельности труда.

Но- мер шах- ты	Абсолют- ная оцен- ка, т/ч	Относи- тельная оценка, %
1	104,8	102,7
2	107,3	105,3
3	106,9	104,2
4	146,8	143,8
5	104,1	102,0
6	109,5	107,4
7	101,9	100,0
8	126,4	123,0
9	131,1	129,8

лесной и деревообрабатывающей промышленности в самостоятельную оптимизируемую систему и моделирование протекающих в ней процессов.

Исследование сложившихся особенностей лесозаготовительного производства края позволило выделить следующие проблемы его развития и размещения на перспективу:

а) рациональную увязку производства и потребления древесины внутри края с учетом вывоза за его пределы и при наличии большого влияния транспортного фактора;

б) совмещение процессов выбытия и наращивания производственных мощностей отдельных лесозаготовительных предприятий исходя из общей потребности в древесине и состояния лесосырьевой базы;

в) учет особенностей функционирования каждого лесозаготовительного предприятия, предполагающий изучение процесса формирования его технико-экономических показателей в зависимости от характеристик эксплуатируемой сырьевой базы, уровня производства, продолжительности периода его функционирования и т. д.

7.16. Учитывая специфику перечисленных проблем, а также накопленный опыт моделирования отрасли, можно выделить три самостоятельных уровня моделирования в рамках решения задачи оптимального перспективного планирования отрасли лесозаготовок с привязкой (уже традиционно сложившейся) к ним моделей различных типов: линейно-программных и статистических.

Была осуществлена постановка задачи оптимизации развития и размещения отрасли лесозаготовок в крае в целом и ее последующая реализация на динамической транспортно-производственной модели. Затем проводилась последующая детализация производственной части моделей, предварительно выделенных территориальных группировок лесозаготовительных предприятий. Здесь основной упор был сделан на оптимизацию процесса сочетания выбывания и наращивания производственных мощностей леспромхозов при отказе от транспортного блока. Особенности функционирования отдельных лесозаготовительных предприятий исследовались при помощи экономико-статистических моделей. Моделировались показатели: себестоимость заготовки 1 м³ древесины, объем вывозки, комплексная выработка и т. д. В конечном счете были сформулированы основные принципы и подходы к построению динамической статистической модели лесоза-

готовительного предприятия (Распознавание образов, 1972).

Указанные постановки моделей имели локальный характер, их цели представляли собой некоторые «вычленения» из общего круга проблем. Комплексной же реализации перечисленных выше проблем на моделях математического программирования не было, а именно она, собственно говоря, означает действительную оптимизацию развития названной отрасли. Поэтому естественным было стремление на следующем шаге увязать уже работающие модели в единую систему оптимизации.

7.17. Структура системы такова. На верхнем уровне (промышленность в целом) решается оптимальная динамическая транспортно-производственная задача, в которой определяются объемы заготовок за плановый период и планы перевозок древесины выделенным потребителям. Дополнительная процедура позволяет выявить наиболее рациональные (с точки зрения замкнутости) территориальные группировки производителей и потребителей древесины. На среднем уровне (территориальная группировка) с помощью оптимальных динамических производственных моделей определяются тенденции выбытия и наращивания производственных мощностей отдельных лесозаготовительных предприятий в планируемом периоде. Наконец, на нижнем уровне (отдельное предприятие) осуществляется статистическое моделирование показателей, характеризующих эффективность производственно-хозяйственной деятельности леспромхоза, в условиях плана территориальной группировки и с учетом специфических условий деятельности предприятия.

Увязка моделей показана на схеме 4, где верхний, средний и нижний уровни системы обозначены соответственно через IA, IB, IB.

7.18. Поясним существо процесса увязки моделей. «Внешними» по отношению к системе параметрами выступают: потребность в древесине j -го пункта потребления в году t — b_{jt} ; доля производства древесины в году t в общем объеме заготовок i -го леспромхоза — α_{it} ; эксплуатационные затраты i -го леспромхоза в году t — c_{it} .

Реализация схемы начинается с решения транспортно-производственной модели. На основе двойственных переменных U_{it} выполняется группировка производственных объектов для следующего уровня моделирования. «Внешним» по отношению к нему параметром будет потребность



Схема 4. Информационный обмен в системе моделей оптимизации лесозаготовительной промышленности Красноярского края.

в древесине по годам планового периода для каждой группировки — b_{ik} , определяемая из плана задачи IA.

Из решения по модели IB получают объемы заготовок древесины для каждого i -го лесопромхоза, входящего в состав K -й территориальной группировки, по годам планового периода — $x_{i1} + \sum_{t'=2}^T \Delta x_{it'}$, которые направляются на нижний уровень моделирования.

На уровне i -го лесопромхоза рассчитываются эксплуатационные затраты для каждого года t планового периода — c_{it} с учетом плана территориальной группировки и специфических условий деятельности предприятия.

С реакции предприятия на полученный план начинается новый поток — «снизу—вверх». В нашем случае реакция отражается в показателе себестоимости заготовки кубометра древесины по годам планируемого периода. Этот показатель посылается на средний уровень, где на его основе рассчитываются коэффициенты целевой функции для используемых там моделей. Новое решение моделей IB посылается на нижний уровень (модель IB). Процесс взаимного информационного обмена между моделями IB и IV заканчивается, когда достигается соответствие эксплуатационных затрат в лесопромхозе распределенным объемам заготовок древесины. Полученные на последнем шаге этого процесса объемы заготовок и эксплуатационные затраты передаются на верхний уровень (модель IA) в виде годовых долей производства в планируемом периоде — α_{it} и значений c_{it} . Начинается новая итерация в схеме информационного обмена моделей.

7.19. При подготовке исходной информации для построения экономико-статистической модели производственного объекта вся совокупность лесопромхозов, действующих на территории края, в соответствии с принятой гипотезой об их поэтапном развитии (см. гл. 6, § 3) была разбита на три группы: развивающиеся лесопромхозы — 25, стабилизировавшиеся в своем развитии — 17 и затухающие — 15. В качестве единичного наблюдения выступал «лесопромхозо-год». Общее число наблюдений по выделенным группам соответственно составляло 126, 112 и 115.

Вид ЭСМ приведен в табл. 7.7.

7.20. Первая итерация модели IA. В результате решения задачи верхнего уровня был получен оптимальный план производства и транспортировки древесины по годам планового периода. Функционал модели IA на 1-й итерации — 578070 тыс. руб. После этого посредством дополнительной процедуры вся совокупность производителей древесины была разбита на 5 групп, исходя из их территориальной принадлежности и оптимальной привязки к потребителям древесины.

В первую группу (по закрепленным в оптимальном плане потребителям назовем ее условно Маклаковской

Экономико-статистические модели объектов

Этап	Масштаб коэффициентов	Коэффициент множественной корреляции	Значения коэффициентов при факторах в регрессионных уравнениях										
			x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7			
I	Стандартизованный	0,947											
	Натуральный		1,41	0,776	-0,31	0,165	0,116	0,09	-0,01	-0,009			
	Стандартизованный	0,901		0,9	-0,005	0,0002	0,001	0,87	-0,001	-0,0004			
II	Натуральный		3,25	0,746	-0,238	0,407	0,208	-0,473	-0,015	-0,124			
	Стандартизованный	0,863		0,75	-0,004	0,00006	0,001	-0,83м	-0,0008	-0,004			
III	Натуральный		2,31	0,72	-0,068	0,083	0,486	-0,293	0,055	-0,117			
	Стандартизованный			0,83	-0,0002	0,0002	0,001	-3,5	0,004	-0,006			

Здесь X_0 — свободный член уравнения регрессии;

X_1 — себестоимость 1 м³ древесины предыдущего года (руб./м³);

X_2 — объем заготовок древесины (тыс. м³);

X_3 — величина промышленно-производственных фондов (тыс. руб.);

X_4 — численность промышленно-производственного персонала (чел.);

X_5 — средний объем хлыста (м³);

X_6 — среднее расстояние вывозки (км);

X_7 — прирост объема заготовок к предыдущему году (тыс. м³).

Примечание. Номера этапов соответствуют трем группам леспромхозов.

и из этих же принципов будем называть и следующие группы) вошли 18 леспромхозов, во вторую, Канскую — 11, в третью, Боготольскую — 11, в четвертую, Красноярскую — 11, наконец, в пятую, Амзасскую — 9 леспромхозов.

7.21. Первая итерация модели IB. По оптимальным моделям территориальных группировок осуществлялось распределение плановых заданий каждому леспромхозу, входящему в их состав, по годам рассматриваемого периода при сохранении суммарной потребности, полученной из модели верхнего уровня. На основе объемов заготовок рассчитывались эксплуатационные затраты в моделях нижнего уровня, а распределение плановых заданий внутри периода позволило сформировать новую структуру производства (z_{it}) для модели верхнего уровня.

Значения функционалов пяти производственных моделей были следующие (в тыс. руб.):

Модель 1	194 324
Модель 2	107 171
Модель 3	89 770
Модель 4	70 247
Модель 5	44 769
Итого	506 281

7.22. Первая итерация модели IB. Эксплуатационные затраты рассчитывались по статистическим моделям. Средняя ошибка прогноза составила 4%.

Далее согласно схеме взаимного информационного обмена в трехуровневой системе следует процесс координации моделей IB и IB. Предварительно, исходя из планов лесозаготовительных предприятий, полученных из решения оптимальных производственных моделей, определялись объемы промышленно-производственных фондов и численность занятых. За основу были взяты коэффициенты фондоемкости и трудоемкости продукции лесозаготовок в 1972 г. для леспромхозов каждого этапа. Показатели «средний объем хлыста» и «среднее расстояние вывозки» формировались следующим образом. В качестве базового использовалось их значение в 1972 г. В последующие 4 года предполагалось уменьшение среднего объема хлыста для леспромхозов I этапа на 0,01 м³, II — на 0,02 м³ и III — на 0,035 м³ в год. Рост среднего расстояния вывозки составлял соответственно: 1,0; 2,0 и 3,5 км в год.

Получив значения всех независимых переменных моделей *IB*, нетрудно было рассчитать величины удельной себестоимости для каждого из лесозаготовительных предприятий по годам планового периода. Трансформировав их в соответствии с процедурой, получили значения коэффициентов целевой функции для моделей *IB*: началась вторая итерация в системе моделей *IB* и *IV*.

7.23. Распределение плановых заданий, найденное из решения моделей *IB* на второй итерации, резко отличается от предшествующего. В общей сложности 33 лесозаготовительных предприятия изменили планы. Получены следующие значения функционалов (в тыс. руб.) на 2-й итерации:

Модель 1	187 876
Модель 2	104 738
Модель 3	86 172
Модель 4	69 528
Модель 5	44 006
Итого	492 320

По аналогии с первой итерацией рассчитывались удельные себестоимости (реализация моделей *IB*) и коэффициенты целевой функции моделей *IB*. Затем следовала третья итерация в системе моделей *IB* и *IV*. На третьей и четвертой итерациях получены такие значения функционалов (в тыс. руб.):

	3-я итерация	4-я итерация
Модель 1	181 502	180 770
Модель 2	104 738	104 467
Модель 3	84 035	83 890
Модель 4	68 593	68 593
Модель 5	44 006	44 006
Итого	482 874	481 726

В итоге четырех итераций сумма функционалов в пяти задачах уменьшилась с 506 282 до 481 726 тыс. руб., т. е. на 24 556 тыс. руб., или примерно на 5%. После этого было решено процесс информационного обмена в системе моделей *IB* и *IV* прекратить и начать формирование новой погодовой структуры заготовки древесины и коэффициентов целевой функции для модели *IA*. Начался второй итерационный цикл в трехуровневой системе моделей *IA*, *IB* и *IV*.

Вторичная реализация транспортно-производственной модели дала функционал в размере 551 432 тыс. руб. По

сравнению с предыдущим значением он уменьшился на 26 638 тыс. руб. (578 070—551 432), или примерно на 4,7%. Далее вновь осуществлялась процедура группировки, обеспечившая новый состав групп: 1—25 лесозаготовительных предприятий; 2—15; 3—8; 4—7 и 5—6 лесозаготовительных предприятий. Начался итерационный процесс в системе моделей *IB* и *IV*. Так как его основные принципы были подробно пояснены ранее, мы заканчиваем изложение.

7.24. Найденный в результате итеративного процесса согласования моделей разных уровней оптимальный план учитывает основные закономерности и особенности, присущие процессу перспективного развития отрасли в целом и в разрезе отдельных лесозаготовительных предприятий. В зависимости от целей исследования настоящая схема без труда упрощается в двухуровневые: *IA—IB*; *IA—IV*; *IB—IV*. С технической стороны реализация предложенной схемы увязки моделей предполагает лишь наличие нескольких дополнительных машинных подпрограмм.

Строгого доказательства сходимости изложенной процедуры согласования не имеется.

ИССЛЕДОВАНИЕ
ВЫБОРОЧНОЙ СОВОКУПНОСТИ ОБЪЕКТОВ
НА ОДНОРОДНОСТЬ

В гл. 3, § 2 данной работы рассматривалась проблема исследования выборочной совокупности на однородность. Ниже приводятся некоторые алгоритмы (программы), позволяющие проводить подобный анализ.

АЛГОРИТМ ВЫДЕЛЕНИЯ
ОДНОРОДНЫХ СОВОКУПНОСТЕЙ ОБЪЕКТОВ

1.1. Алгоритм¹ относится к классу статистических и применим для многомерной классификации. Один из подходов к решению задачи классификации состоит в том, чтобы отобразить выборочное пространство в некоторое другое, где бы свойство неоднородности значений меры близости между любой парой наблюдений смешанной выборки отобразилось в иное свойство, более удобное для целей классификации. Если в качестве пространства отображений принять функцию плотности вероятности смеси распределений, — а так как она неизвестна, то ее оценку, вычисленную каким-либо способом для любой точки выборочного пространства, — то, очевидно, выборочное пространство перейдет во множество значений оценки функции плотности вероятности смеси. Неравномерное распределение объектов в пространстве отобразится в много-

¹ Беккер А. В., Ягольницер М. А. Алгоритм выделения однородных совокупностей объектов. — В кн.: Распознавание образов и регрессионный анализ в экономических исследованиях. Новосибирск, 1972, с. 95—105.

дальное изменение оценки функции плотности вероятности смеси. Каждая область «сгущений» отображается в область значений плотности, тяготеющих к одной и только одной ее моде, а границы между «сгущениями» — в «овраги» функции плотности вероятности смеси.

1.2. Алгоритм использует подход, связанный с непараметрической оценкой плотности вероятности по имеющейся выборке. Будем считать, что вершины совместной плотности распределения соответствуют «центрам» групп, а поверхностям, проходящие в долинах этого рельефа, — границам групп; количество существенных максимумов функции плотности вероятности определяет число групп. Алгоритм выделяет вершины (объекты, образующие эти вершины) в многомерном распределении, построенном по имеющейся выборке.

Исходными данными выступают предъявленные для группировки точки

$$Q^0 = \{x_1^0, x_2^0, \dots, x_n^0\},$$

где $x_j^0 = \{x_{j1}^0, x_{j2}^0, \dots, x_{jm}^0\}$, $j = 1, 2, \dots, n$;

m — размерность признакового пространства.

Алгоритм представляет собой итерационную процедуру, которая по точкам $Q^t = \{x_1^t, x_2^t, \dots, x_n^t\}$ определяет новые точки $Q^{t+1} = \{x_1^{t+1}, x_2^{t+1}, \dots, x_n^{t+1}\}$, более отчетливо группирующиеся в таксоны, чем точки в выборке Q^t . Каждая итерация содержит шесть этапов.

Этап 1. Вычисление ковариационной матрицы связи признаков в выборке Q^t по формуле

$$\Sigma_t = \sum_{j=1}^n p_j^t \left(x_j^t - \sum_{i=1}^n p_i^t x_i^t \right) \left(x_j^t - \sum_{i=1}^n p_i^t x_i^t \right)',$$

где Σ_t — выборочная ковариационная матрица размерности $m \cdot m$ на t -й итерации²;

p_j^t — вероятность появления j -го таксона.

Этап 2. Вычисление меры связи между объектами по формуле

$$\mu(x_j^t, x_k^t) = \exp \left\{ -\frac{1}{2h^2} (x_j^t - x_k^t)' \Sigma_t^{-1} (x_j^t - x_k^t) \right\},$$

² В случае независимых признаков ковариационная матрица переходит в диагональную матрицу $m \cdot m$, элементы которой представляют дисперсии соответствующих признаков.

где h — параметр, характеризующий точность аппроксимации функции плотности вероятности.

Этап 3. Вычисление вероятности α_{ik}^t принадлежности объекта i к некоторому таксону k по формуле

$$\alpha_{ik}^t = \frac{p_k^t \mu(x_i^t, x_k^t)}{\sum_{j=1}^n p_j^t \mu(x_i^t, x_j^t)}$$

Этап 4. Вычисление вероятности p_k^{t+1} , $k=1, 2, \dots, n$ по формуле

$$p_k^{t+1} = \frac{1}{n} \sum_{i=1}^n \alpha_{ik}^t$$

Этап 5. Определение выборки

$Q^{t+1} = \{x_1^{t+1}, x_2^{t+1}, \dots, x_n^{t+1}\}$ по формуле

$$x_k^{t+1} = \sum_{i=1}^n \alpha_{ik}^t x_i^t / \sum_{i=1}^n \alpha_{ik}^t, \quad k=1, 2, \dots, n.$$

Этап 6. Проверка критерия прекращения итерационного процесса вида

$$\max_i (x_i^t - x_i^{t+1}) \sum_i^{-1} (x_i^t - x_i^{t+1})' \leq \epsilon,$$

где $\epsilon > 0$ — заданное число.

Для выбранного значения параметра h число таксонов $1 \leq k \leq n$ получается автоматически. В один таксон попадают объекты, характеристики которых с некоторой точностью можно считать одинаковыми.

1.3. Алгоритм «Рельеф» рассчитан на классификацию объектов, описанных набором количественных признаков. Отнесение объектов к соответствующим классам происходит на основании подсчета вероятностей принадлежности объектов к получаемым классам.

Классификация определяется значением параметра аппроксимации функции плотности вероятности (управляющего параметра алгоритма). Чем лучше выбрано значение h для аппроксимации плотности вероятности, тем выше качество классификации. Число классов заранее не задается, а получается автоматически.

По данному алгоритму написана программа³ на языке Алгол (входной язык α -транслятора). Работа программы обеспечивает выдачу результатов классификации (для заданного управляющего параметра), а также матрицы вероятностей принадлежности объектов к полученным классам.

ДРЕВОВИДНАЯ ГРУППИРОВКА НА ОСНОВЕ КАЧЕСТВЕННЫХ ПРИЗНАКОВ

1.4. Программа⁴ реализует алгоритм комплексного решения четырех задач, связанных общей целью — выделением из совокупности изучаемых объектов качественно однородных групп. Специфика задач определяется различием требований, предъявляемых к группировке. Области применения этих задач являются анализ статистической совокупности на однородность, построение дискретных и дискретно-непрерывных моделей (см. гл. 3, § 2—3). При разработке алгоритма использовался метод последовательного разбиения совокупности по качественным признакам⁵.

1.5. **Задача 1 — группировка по формальному критерию.** Имеется статистическая совокупность объектов, каждый из которых описывается набором качественных признаков. Необходимо подразделить совокупность на ограниченное число групп похожих объектов так, чтобы группы максимально отличались друг от друга. Другими словами, задача состоит в том, чтобы при заданном уровне дробности совокупности получить наиболее однородные в качественном плане группы. Ограничение на число групп обычно связано с необходимостью обеспечить надежность статистических оценок.

³ Ягольницер М. А. Программа таксономии «Рельеф». — В кн.: Алгоритмы статистической обработки информации. Новосибирск, 1974, с. 80—87.

⁴ Журавель Н. М., Журавель Ф. А. Последовательная группировка на основе качественных признаков. — В кн.: Распознавание образов и регрессионный анализ в экономических исследованиях. Новосибирск, 1972, с. 45—59.

⁵ Розин Б. Б. Учет влияния качественных признаков при моделировании экономических показателей. — В кн.: Вопросы экономико-статистического моделирования и прогнозирования в промышленности. Новосибирск, 1970, с. 168—201.

Следовательно, в данном случае ставится некоторая экстремальная задача: из всевозможных группировок по n групп, которые можно получить, варьируя группировочные признаки, найти такую, чтобы совокупное число различий объектов по группам было максимальным.

Может быть поставлена и обратная задача: определить уровень дробности совокупности, который бы в достаточной мере обеспечивал однородность формируемых групп.

Алгоритм решения этих задач представляет многошаговую процедуру: на первом шаге полная совокупность объектов разбивается по какому-либо одному признаку, на втором шаге каждая из полученных групп разбивается по своему признаку и т. д.

Главный вопрос организации подобного процесса состоит в выборе группировочного признака для каждого частного разбиения. Здесь применимы как априорные оценки важности признаков, выдвигаемых на основе профессиональных знаний, так и формальные критерии. Формальным критерием при выборе группировочного признака может служить величина коэффициента качественной вариации Q (Мюллер, Шусслер, 1968). Расчет этого коэффициента строится на выявлении полного числа различий в качествах всех объектов совокупности. Это число определяется суммированием произведений частот каждой градации признака на частоту каждой отличной от нее градации. Взяв отношение полного числа различий к максимально возможному (когда все частоты различных градаций равны), получим величину Q . Для случая нескольких градаций качественного признака формула расчета Q имеет вид

$$Q = \frac{n_1 \sum_{i=2}^P n_i + n_2 \sum_{i=3}^P n_i + \dots + n_{p-2} \sum_{i=p-1}^P n_i + n_{p-1} \cdot n_p}{p(p-1)/2(N/P)^2}, \quad (1)$$

где P — число градаций качественного признака;

n_i — частота, с какой встречается в совокупности i -я градация ($i=1, \dots, P$);

N — численность всей совокупности.

Прямая задача может быть решена, если при заданном числе шагов разбиение на каждом шаге осуществляется по признаку, для которого значение коэффициента качественной вариации Q максимально. Решение обратной задачи находится путем анализа уменьшения численности

групп при увеличении числа шагов в процедуре последовательных разбиений. В данном типе задач коэффициент качественной вариации Q используется без ограничений.

1.6. Задача 2 — группировка с учетом априорных оценок важности признаков. Как и в первой задаче, разбивается совокупность объектов, по каждому из которых есть набор качественных признаков. Но назначение группировки предусматривает использование выделенных качественно однородных групп для определенной практической цели. Например, в случае разбиения совокупности промышленных объектов это может быть создание устойчивой пространственно-временной типологии, выявление уровней технического прогресса и т. п. При разбиении совокупности социальных объектов внешняя цель может состоять, например, в изучении степени мобильности различных социальных групп, в выявлении законов распределений свободного времени и т. п.

В задачах такого рода весьма желательна дополнительная информация о содержательной ценности качественных признаков по отношению к внешней цели. Здесь важно, чтобы при разбиении выбор группировочного признака осуществлялся с корректировкой критерия однородности на степень ценности признака.

Введение такой информации в процедуру разбиений можно осуществить с помощью следующего методического приема. Все группировочные признаки ранжируются по степени ценности с позиций внешней цели. Эта ранжировка используется как шкала предпочтений признаков. Группировочный признак на каждом шаге выбирается по шкале предпочтений из числа признаков, величина Q для которых лежит в интервале $(Q_{\max} \pm \epsilon)$. Меняя величину ϵ , можно регулировать при выборе группировочного признака глубину связи между формальным критерием и шкалой предпочтений.

1.7. Задача 3 — группировка под ступенчатую и кусочно-линейную аппроксимацию. Подобные задачи возникают при математико-статистическом моделировании каких-либо результирующих показателей функционирования объектов совокупности (например, технико-экономических показателей — для промышленных объектов, показателей урожайности, выживаемости и т. п. — для биологических объектов).

Здесь специфика задачи разбиения заключается в необходимости формировать группы по тем признакам, кото-

рые в большей мере влияют на моделируемый показатель. Для этой цели при выборе группировочного признака коэффициенты качественной вариации нужно скорректировать на величину коэффициентов K — связи признаков с моделируемым показателем, например простым перемножением $Q \times K$.

В качестве коэффициентов K могут быть применены коэффициенты сопряженности или корреляций в зависимости от того, качественную или количественную природу имеют признаки. Естественно, что вид коэффициентов, используемых в конкретной задаче, должен быть одинаков для всех признаков. Следовательно, если среди признаков есть хоть один качественный, то используются только коэффициенты сопряженности. Для расчета последних по количественным признакам необходим искусственный перевод их в качественные путем выделения характерных интервалов варьирования. Ограничение на вид коэффициентов вызвано тем, что рассчитанные для одной и той же пары количественных признаков, коэффициенты сопряженности и корреляции численно не совпадают, первые по природе своей занижают степень связи.

1.8. Задача 4 — разбиение по заданной схеме. Такая задача является своего рода дополнением к любой из описанных ранее. Необходимость в ней возникает, когда в целях сравнительного анализа известную уже схему разбиения требуется повторить для другой совокупности объектов. Поскольку при этом наименования группировочных признаков, используемых на всех шагах, заранее задаются, то изменения могут наблюдаться лишь в составе получаемых групп.

Сравнительный анализ может проводиться в двух разрезах: пространственном и временном. В первом случае разбиение выполняется на основе информации за один и тот же период по совокупностям, различающимся, например, административным подчинением, географическим положением и т. п. В результате анализа выявляются структурные закономерности изучаемого явления.

При временном разрезе анализируются разбиения одной и той же совокупности объектов за ряд последовательных периодов. При их сравнении вскрываются закономерности динамики формируемых групп.

Возможна комбинация пространственного и временного разрезов. Для этого необходима информация о всех

пространственных совокупностях за исследуемые периоды. Результаты сравнительных анализов такого рода, очевидно, могут быть использованы как для того, чтобы вскрыть механизм изучаемого явления, так и для прогноза его развития на будущее.

1.9. Исходная информация для машинной реализации упомянутых задач задается в виде двухмерного массива $A[o:m, o:n]$, где m — количество объектов, а n — количество признаков. Значение элемента $A[i, j]$ соответствует померу градации j -го признака для i -го объекта. В нулевом столбце расположены номера объектов, а в нулевой строке — максимальное количество градаций по данному признаку.

На первом шаге разбиения формируется первый вектор номеров объектов, соответствующий нулевому столбцу исходной матрицы, и вектор границ подмассивов. Очевидно, что на первом шаге номера объектов расположены в естественном порядке возрастания, а вектор границ подмассивов содержит одно значение, равное количеству классифицируемых объектов.

Затем по формуле (1) рассчитываются значения коэффициента для каждого признака. В зависимости от типа задачи, который задается информационной картой при вводе числовых данных, выбирается признак разбиения. По выбранному признаку определяется количество подмассивов разбиения, равное числу градаций этого признака, и подсчитывается длина каждого подмассива. Затем поочередно номера объектов из первого вектора размещаются согласно выделенным подмассивам, формируя второй вектор объектов. При этом номер подмассива для каждого объекта определяется его принадлежностью к той или иной градации признака, по которому выполняется разбиение. Далее следует печать коэффициентов однородности Q , количества объектов и их номеров в каждом подмножестве. Наконец, пересчитывается вектор границ подмассивов, исходя из количества объектов в них, и в первый вектор номеров объектов переписываются значения второго вектора. Описанный алгоритм повторяется для всех последующих шагов разбиения, при этом полученные на предыдущем шаге подмассивы рассматриваются последовательно. Если на каком-то шаге получается нулевой подмассив, то при дальнейшем разбиении он игнорируется и информация о нем на печать не выводится.

ГРУППИРОВКА ЭЛЕМЕНТОВ МЕТОДОМ ДИАГОНАЛИЗАЦИИ МАТРИЦЫ РАССТОЯНИЙ

1.10. Программа⁶ позволяет по матрице взаимных расстояний (или матрице мер близости) разбить элементы x_1, x_2, \dots, x_N на группы таким образом, чтобы в одну группу попали наиболее близкие элементы. Основная идея используемого алгоритма состоит в построении последовательности элементов $x_{i_1}, x_{i_2}, \dots, x_{i_N}$, для которой $x_{i_v}, x_{i_{v+1}} \in M \Rightarrow \Rightarrow x_{i_{v+1}}, x_{i_{v+2}}, \dots, x_{i_{v+r-1}} \in M$. Соответственно матрица расстояний приобретает блочно-диагональный вид.

1.11. Программа может быть использована при проведении такого важного этапа экономико-статистического исследования, как анализ совокупности объектов на однородность и, в случае выявления неоднородности изучаемой совокупности, разбиение объектов на однородные группы (гл. 3, § 2). В этом случае анализируется матрица расстояний $\|r_{ij}\|$, $i, j=1, 2, \dots, N$, где r_{ij} — расстояние между объектами x_i и x_j .

Аналогичные проблемы, только по отношению не к самим объектам, а к их признакам, возникают также в процессе построения регрессионных моделей (гл. 2, § 4, гл. 3, § 2). Чтобы решить, какие именно признаки и в каком количестве необходимо включить в модель, надо определить группы признаков, тесно коррелированных между собой; группы признаков, примерно одинаково связанных с признаком-показателем, для которого строится регрессионная модель; группы признаков, которые в первом приближении можно считать независимыми, и пр. В этом случае анализируется матрица мер близости признаков.

Анализ матрицы взаимных расстояний исследуемой совокупности объектов и матрицы мер близости (коэффициентов связи — корреляции, ковариации и пр.) признаков объектов позволяет решить некоторые из этих задач. Поэтому и объекты совокупности, и их признаки будем называть «элементами статистической совокупности» и анализировать структуру соответствующей им матрицы взаимных расстояний, или коэффициентов близости.

⁶ Говорская Л. В. Группировка элементов методом диагонализации матрицы расстояний. — В кн.: Алгоритмы статистической обработки информации. Новосибирск, 1974, с. 55—72.

1.12. Используемый в программе алгоритм диагонализации матрицы⁷ привлекателен тем, что не требуется заранее знать количество групп, на которые распадается исследуемая совокупность, так как эти группы получаются автоматически в процессе применения алгоритма.

Будем считать матрицу взаимных расстояний (или матрицу мер близости) элементов изучаемой совокупности заданной. В программе предусмотрены четыре критерия, по которым элементы $x_{j_1}, x_{j_2}, \dots, x_{j_N}$ располагаются в последовательность $x_{i_1}, x_{i_2}, \dots, x_{i_N}$, обладающую тем свойством, что если два элемента последовательности принадлежат одному классу (группе), то все элементы последовательности, находящиеся между ними, принадлежат тому же классу.

Введем следующие обозначения:

I — множество индексов $\{1, 2, \dots, N\}$;

I_k — множество индексов $\{i_1, i_2, \dots, i_k\}$, где $i_j \leq N$, $j=1, 2, \dots, k$;

V_k — дополнение I_k к $I(I/I_k)$.

А. Критерий минимального расстояния применяется для анализа матрицы взаимных расстояний $R = \|r_{ij}\|$, $r_{ij} = r_{ji} \geq 0$, $r_{ii} = 0$, $i, j=1, 2, \dots, N$.

На первом этапе выбирается некоторый элемент x_{i_1} из заданной совокупности $x_{j_1}, x_{j_2}, \dots, x_{j_N}$. На втором по матрице взаимных расстояний R находим элемент x_{i_2} такой, что

$$r_{i_1 i_2} = \min_{\substack{j \in I \\ j \neq i_1}} r_{i_1 j}.$$

Если найдены первые K элементов последовательности $x_{i_1}, x_{i_2}, \dots, x_{i_k}$, то следующий элемент $x_{i_{k+1}}$ выбирается так, чтобы выполнялось соотношение

$$r_{d i_{k+1}} = \min_{\substack{i \in I_k \\ v \in V_k}} r_{iv}, \text{ где } d \in I_k.$$

Иными словами, выбирается такой элемент из оставшихся $N-k$, который наиболее близок к одному из K элементов,

⁷ Беккер А. В., Лукацкая М. Л. Об анализе структуры матрицы коэффициентов связи. — В кн.: Вопросы экономико-статистического моделирования и прогнозирования в промышленности. Новосибирск, 1970, с. 273—284.

уже включенных в последовательность. В результате элементы совокупности записываются в последовательность $x_{i_1}, x_{i_2}, \dots, x_{i_N}$ по группам.

В. Критерий минимального среднего расстояния. Программа позволяет для той же матрицы взаимных расстояний проводить расположение в последовательность и по другому критерию — минимуму среднего расстояния. В этом случае $(k+1)$ -й элемент выбирается следующим образом:

$$\sum_{i \in I_k} r_{ii_{k+1}} = \min_{v \in V_k, i \in I_k} \sum r_{iv}.$$

Иными словами, выбирается такой элемент из оставшихся $N-k$, для которого среднее расстояние до k уже включенных в последовательность элементов будет минимальным. Этот критерий целесообразно применять, если исследуемая матрица не удовлетворяет всем требованиям матрицы расстояний.

Если R есть матрица мер близости (например, матрица коэффициентов корреляции)

$$R = \|r_{ij}\|, r_{ij} = r_{ji}, r_{ii} = 1, i, j = 1, 2, \dots, N,$$

то для нее тоже возможны два критерия.

С. Критерий максимума близости. При расположении элементов $x_{j_1}, x_{j_2}, \dots, x_{j_N}$ в последовательность по критерию C в алгоритме, изложенном выше (критерий A), \min заменяется на \max и берется модуль соответствующих элементов. Таким образом, формулы примут вид:

$$r_{i_1 i_2} = \max_{\substack{j \in I \\ j \neq i_1}} |r_{i_1 j}|;$$

$$r_{di_{k+1}} = \max_{\substack{i \in I_k \\ v \in V_k}} |r_{iv}|, \text{ где } d \in I_k.$$

Д. Критерий максимума средней близости применим к матрице мер близости и аналогичен критерию B . Первая формула в критерии C остается прежней, а вторая заменяется на

$$\sum_{i \in I_k} r_{ii_{k+1}} = \max_{v \in V_k} \sum_{i \in I_k} |r_{iv}|.$$

Таким образом, программа позволяет по одному из четырех критериев найти последовательность $x_{i_1}, x_{i_2}, \dots, x_{i_N}$

и получить преобразованную матрицу \tilde{R} , соответствующую этой последовательности.

1.13. Если задано «пороговое значение» критерия, то программа не только располагает элементы в последовательность, но и автоматически получает их группировку. Тогда алгоритм немного видоизменяется. Например, для критерия A не только отыскивается элемент с минимальным расстоянием ко всем предшествующим элементам данной группы, но это расстояние сравнивается с пороговым значением ρ , и если оно меньше или равно ρ , соответствующий элемент включается в группу, если оно больше, начинается формирование следующей группы.

Пороговое значение критерия является в этом случае некоторой характеристикой компактности образующихся групп. Например, если для первого критерия порогом задается величина ρ , то в одну группу входят только те элементы, для которых минимальное расстояние от других элементов данной группы не превышает ρ . (Если исследователю не нужна автоматическая группировка, то для первых двух критериев (A и B) в качестве ρ необходимо задать число, заведомо большее любого элемента исходной матрицы расстояний; для последних двух критериев (C и D) необходимо ρ взять равными любому отрицательному числу.)

ИССЛЕДОВАНИЕ КАЧЕСТВА КЛАССИФИКАЦИИ

В данном приложении рассмотрены два критерия, позволяющие судить о качестве получаемых с помощью алгоритма классификации «Рельеф» (приложение 1) разбиений объектов на однородные группы.

ИНФОРМАЦИОННЫЙ СТАТИСТИЧЕСКИЙ КРИТЕРИЙ
ПРОВЕРКИ ГИПОТЕЗЫ
ОБ ОДНОРОДНОСТИ СРЕДНИХ ЗНАЧЕНИЙ
И КОРРЕЛЯЦИОННЫХ МАТРИЦ

Критерий предложен Л. А. Верховской (Верховская, 1971). Он основан на проверке статистической гипотезы об однородности средних значений и корреляционных матриц нескольких независимых нормальных многомерных выборок. Применение данного критерия к вышеуказанному алгоритму «Рельеф» обусловлено тем, что аппроксимация функции плотности вероятности в этом алгоритме производится нормальными ядрами.

Пусть имеется d независимых классов по n_i ($i = \overline{1, d}$) независимых объектов в каждом m -мерного признакового пространства с неизвестными математическими ожиданиями и корреляционными матрицами.

Проверяется гипотеза H_0 : векторы всех средних значений и корреляционных матриц всех d классов равны. Конкурирующая альтернативная гипотеза H_1 формулируется так: не все изучаемые параметры равны, т. е.

$$H_0: m_1 = m_2 = \dots = m_d; \quad k_1 = k_2 = \dots = k_d;$$

$$H_1: m_1, m_2, \dots, m_d; \quad k_1, k_2, \dots, k_d.$$

В основу построения критерия положена статистика минимума различающей информации $I(*:H)$. Она яв-

ляется мерой направленного расхождения между выборкой и гипотетическим распределением в случае проверки простой гипотезы, а в случае проверки сложной гипотезы — между выборкой и ближайшим членом из семейства распределений H . Критическая область описывается выражением

$I(*:H_0) - I(*:H_1) \geq c$, где c — пороговая константа, с помощью которой можно регулировать величину ошибки 1-го рода α .

$$P\{I(*:H_0) - I(*:H_1) \geq c/H_0\} \leq \alpha.$$

Для вычисления информационных статистик необходимо вычислить оценку вектора средних и выборочную ковариационную матрицу:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

$$S_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)'$$

После подсчета характеристик определяем информационную статистику для проверки гипотезы об однородности средних значений и корреляционных матриц, имеющей вид:

$$Q_{m,k} = J(*:H_0) - J(*:H_1) = \\ = \frac{1}{2} \sum_{i=1}^d (n_i - 1) \ln(|S_0|/|S_i|),$$

где

$$S_0 = \frac{1}{n-d} \sum_{i=1}^d \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})',$$

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_{ij}.$$

Затем вычислим вторую информационную статистику — для проверки гипотезы об однородности только средних значений при условии равенства ковариационных матриц всех d классов, т. е.

$$H'_0: m_1 = m_2 = \dots = m_d \text{ при } k_1 = k_2 = \dots = k_d,$$

$$H'_1: m_1, m_2, \dots, m_d \text{ при } k_1 = k_2 = \dots = k_d,$$

$$Q_{m,k} = \frac{n-d}{2} \ln(|S_0|/|S|),$$

где

$$S = \frac{1}{n-d} \sum_{i=1}^d (n_i - 1) S_i,$$

$$S_0 = \frac{1}{n-d} \sum_{i=1}^d \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(x_{ij} - \bar{x})'.$$

Аппроксимация распределения информационных статистик осуществляется нецентральным χ^2 -распределением с числом степеней свободы, равным

$$t = (d-1)m(m+3)/2,$$

и параметром нецентральности

$$\beta = \sum_{i=1}^d \left(\frac{1}{n_i - 1} - \frac{1}{n-d} \right) \frac{m(2m^2 + 3m - 1)}{12} + \frac{m(d-1)(m-d+2)}{2(n-d)}.$$

Эта аппроксимация позволяет при сравнении вычисленной статистики с нецентральным χ^2 -распределением при любом уровне значимости принять или отклонить гипотезу об однородности средних значений и корреляционных матриц. Если значение информационной статистики больше табличного значения χ^2 -распределения с t степенями свободы и β -параметром нецентральности при заданном уровне значимости, то данную гипотезу принимаем, в противном случае — отбрасываем.

**ЭВРИСТИЧЕСКИЙ КРИТЕРИЙ
ВЫБОРА ВАРИАНТА КЛАССИФИКАЦИИ
ПО МАТРИЦЕ ВЕРОЯТНОСТЕЙ
ПРИНАДЛЕЖНОСТИ ОБЪЕКТОВ К КЛАССАМ**

Рассматривая матрицу вероятностей принадлежности классифицируемых объектов к полученным классам

$$\|p_{ij}\| \quad (i = \overline{1, n}; j = \overline{1, d}),$$

где n — число объектов, подлежащих классификации,

d — число классов разбиения,

можно предложить достаточно простые эвристические приемы анализа классификаций.

Следует заметить, что элементы матрицы вероятностей принадлежности отражают в определенной степени компактность той или иной классификации. Действительно, при хорошей разделимости классов все объекты с достаточно большой вероятностью (задаваемый порог) принадлежат определенным классам и с незначительной — остальным.

То есть если переставить строки этой матрицы так, что объекты из одного класса будут следовать друг за другом, то матрица примет блочно-диагональный вид. Для совершенно неразделимых классов матрица будет состоять из чисел, равных $1/n$. Противоположный случай — каждый элемент образует самостоятельный класс: матрица единичная (на главной диагонали все единицы, а остальные элементы — нули).

Для удобства работы с матрицей ее представляют в другом виде. Выполним преобразование элементов матрицы по следующему правилу:

$$p_{ij}^* = \sum_{i \in A_i} p_{ij}/n_i, \quad i = \overline{1, d}.$$

В результате такого преобразования от матрицы размерности $n \times d$ переходим к матрице размерности $d \times d$, в которой элементы i -й строки характеризуют долю объектов класса A_i во всех остальных классах A_1, A_2, \dots, A_d . «Изолированность» классов будет тем больше, чем ближе эта матрица к диагональной. Введем такую относительную характеристику матрицы, как коэффициент диагональности. Он вычисляется как частное от деления суммы долей, расположенных на главной диагонали, на общую сумму долей всей матрицы:

$$Q = \sum_{i=1}^d p_{ii} / \sum_{i,j=1}^d p_{ij}.$$

В идеальном случае, т. е. когда классы не пересекаются, $Q = 1$.

Другой характеристикой компактности является средняя величина энтропии, подсчитываемая следующим образом. Введем меру неопределенности каждой строки матрицы $\|p_{ij}^*\|$:

$$H_i = - \sum_{j=1}^d p_{ij}^* \ln p_{ij}^*, \quad i = \overline{1, d}.$$

Тогда средняя величина неопределенности разбиения выражается через следующее соотношение:

$$H = \sum_{i=1}^d p_i H_i,$$

где p_i — априорная вероятность появления класса A_i ($p_i = n_i/n$)

(n_i — число объектов класса A_i),

n — общее число объектов, подлежащих классификации.

Нетрудно заметить, что последняя мера является монотонной характеристикой разбиения. Для полностью разделимых классов $H=0$ и возрастает по мере пересечения классов.

СРАВНЕНИЕ ФОРМАЛЬНОГО И ЭВРИСТИЧЕСКОГО КРИТЕРИЕВ ВЫБОРА ВАРИАНТА КЛАССИФИКАЦИИ

Сравнение двух рассматриваемых критериев позволяет вскрыть преимущества и недостатки каждого в отдельности и определить пути их возможного использования в каждом отдельном случае.

Важным преимуществом формального информационного статистического критерия является возможность нахождения аппроксимации для закона распределения этого критерия. Аппроксимация распределения позволяет определить пороговые константы C , обеспечивающие заданную величину ошибки 1-го рода α .

Недостаток данного критерия заключается в значительной трудоемкости вычисления статистик. Возникают ситуации, когда число элементов в классе мало, что приводит к получению неустойчивых оценок в таких классах (вектора средних и ковариационной матрицы). Это обстоятельство заставляет отбрасывать малочисленные классы (один — три объекта) и проводить анализ только оставшихся классов. На наш взгляд, такой прием не должен существенно повлиять на дальнейшие результаты исследования. Дело в том, что результаты классификации используются в дальнейшем при построении внутриклассовых моделей. Появление малочисленных классов объектов не позволяет строить для них статистические модели.

И поскольку для таких классов построение моделей отличается от основной методики, необходимо проверить существенность различия лишь в «больших» классах объектов.

Эвристический критерий качества разбиения использует матрицу вероятностей принадлежности объектов к классам. Исследуя эту матрицу, можно сделать важные выводы относительно компактности классов, степени их изолированности и т. п.

Относительно простые характеристики, подсчитанные на основе матрицы вероятностей принадлежности, позволяют достаточно корректно определить приемлемое разбиение.

Следует отметить, что характеристики, вычисленные по матрице вероятностей принадлежности объектов к классам, являются, вообще говоря, случайными величинами и поэтому имеют некоторое распределение. Вид этого распределения пока не найден, а значит, невозможно найти сравнительную эффективность информационного и эвристического критериев качества разбиения.

Интуитивно понятно, что эффективность этих критериев должна быть примерно одинаковой, так как оба они основаны на гипотезе компактности.

С вычислительной точки зрения реализация эвристического критерия существенно проще, чем информационного. Кроме того, последний критерий более устойчив к виду распределения выборочной совокупности, поскольку априорно не учитывает вид исходного распределения объектов.

По-видимому, дальнейшее изучение свойств матрицы вероятностей принадлежности поможет получить точное (или приближенное) распределение эвристического критерия.

Необходимо отметить, что если распределение исходных объектов хорошо описывается суперпозицией нормальных распределений и выполняется гипотеза компактности классов, то целесообразно при анализе классификации пользоваться информационным статистическим критерием. Если же классы существенно перекрываются и линейно не разделяются, то следует применять эвристический критерий выбора классификации.

**АЛГОРИТМЫ
КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА**

3.1. При построении экономико-статистических моделей (непрерывных и дискретно-непрерывных, статических и динамических (гл. 3, 4)) можно пользоваться программами (или комплексом программ), позволяющими по исходной информации о работе группы объектов в течение определенного периода времени определить параметры ЭСМ и их статистические оценки. В данном приложении приводятся три алгоритма программ такого рода, получивших наибольшее применение при выполнении наших исследований.

АЛГОРИТМ «РЕГРЕССИЯ»

3.2. Программа «Регрессия»¹ по заданной матрице парных коэффициентов корреляции, векторам средних значений переменных и их дисперсий находит коэффициенты регрессий в стандартизованном и натуральном масштабах, вычисляет важнейшие статистические характеристики: коэффициент детерминации, стандартную ошибку оценки, дисперсионное отношение Фишера, дисперсии коэффициентов регрессии.

3.3. Уравнение множественной линейной регрессии в стандартизованном масштабе имеет вид

$$t_1 = \beta_2 t_2 + \beta_3 t_3 + \dots + \beta_p t_p,$$

¹ Программа написана Л. В. Говорской в кодах М-220 (ИЭиОПП СО АН СССР).

где t_i — стандартизованное значение переменной x_i ($t_i = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$, $i=2, 3, \dots, p$); $\beta_2, \beta_3, \dots, \beta_p$ — стан-

дартизованные коэффициенты множественной регрессии. 3.4. По заданной матрице парных коэффициентов корреляции β -коэффициенты находятся методом наименьших квадратов, который приводит к системе нормальных уравнений:

$$\begin{cases} r_{12} = \beta_2 + \beta_3 r_{32} + \beta_4 r_{42} + \dots + \beta_p r_{p2} \\ r_{13} = \beta_2 r_{23} + \beta_3 + \beta_4 r_{43} + \dots + \beta_p r_{p3} \\ \dots \\ r_{1p} = \beta_2 r_{2p} + \beta_3 r_{3p} + \beta_4 r_{4p} + \dots + \beta_p. \end{cases}$$

3.5. В качестве функции может быть взята любая из n переменных, а аргументом — любая группа из остальных переменных ($2 \leq p-1 \leq n$). Функция и аргументы уравнений задаются исследователем в информационной карте программы. Одновременно может быть проведен расчет параметров 15 моделей для различных функций. Анализ статистических характеристик полученных регрессий, коэффициентов регрессий и их доверительных интервалов позволяет исследователю выбрать наиболее приемлемый вариант уравнения регрессии.

3.6. Матрица парных коэффициентов корреляции может быть получена, в частности, по программе расчета парных коэффициентов корреляции при неограниченном числе наблюдений (Статистические алгоритмы и программы, 1970, с. 127—139) или по программе расчета корреляционной матрицы с учетом весов наблюдений², или по любой другой известной программе вычисления парных коэффициентов корреляции. Отличие второй программы состоит в том, что она используется для вычисления указанных характеристик в совокупностях, предварительно исследованных на однородность, учитывает вероятность принадлежности объектов выборки к соответствующим классам и выдает корреляционные матрицы с учетом этих весов для каждого класса отдельно. Используя в ка-

² Программа написана М. А. Ягольницером (ИЭиОПП СО АН СССР).

честве исходных данных для программы «Регрессия» такие корреляционные матрицы, можно построить дискретно-непрерывную модель (гл. 3, § 4).

3.7. Программа расчета парных коэффициентов корреляции при неограниченном числе наблюдений позволяет проводить различные преобразования первоначальных переменных и рассчитывает корреляционную матрицу по расширенному набору переменных, включающему и первоначальные и преобразованные переменные (например, вместе с переменной x могут быть учтены переменные x^2 , $\ln x$, $1/x$ и т. д.). Использование такой расширенной корреляционной матрицы в программе «Регрессия» позволяет получить также нелинейные регрессии.

ПОСТРОЕНИЕ РЕГРЕССИИ МЕТОДОМ ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

3.8. Рассматриваемая регрессионная программа³ рассчитана на построение множественных регрессий как для равнооточных, так и для неравнооточных наблюдений (взвешенная регрессия). Для получения взвешенной регрессии необходимо задать вектор, длина которого равна количеству наблюдений (объектов), а содержимое компонент определяет меру веса соответствующего объекта.

3.9. Так как программа позволяет одновременно анализировать и строить регрессионные модели для нескольких вектор-весов, то ее целесообразно использовать при моделировании совокупности, обладающей определенной типологической структурой. При этом необходимо знать классификацию исследуемой совокупности (группы однородных объектов) и последовательность вектор-весов, трактуемых в данном случае как вероятности принадлежности объектов заданным классам. (Такая матрица весов принадлежности может быть получена, в частности, по программе таксономии «Рельеф».) Полученная последовательность взвешенных регрессий и будет совокупностью внутриклассовых моделей, определяющей вместе с описанием классов дискретно-непрерывную модель (гл. 3, § 4).

³ Лукацкая М. Л., Ульянова Г. П. Построение взвешенной регрессии методом последовательных приближений. — В кн.: Алгоритмы статистической обработки информации. Новосибирск, 1974, с. 5—26.

3.10. Регрессионная программа может быть использована для анализа временных рядов и построения динамических моделей в том случае, когда наблюдения, соответствующие информации о более отдаленном периоде времени, хотя и вводят в модель с меньшими весами, чем веса наблюдений, непосредственно предшествующих периоду, в котором модель будет использоваться.

3.11. Программа работает в нескольких режимах, что дает возможность для равнооточных и неравнооточных наблюдений как непрерывные, так и дискретно-непрерывные модели строить тремя способами: для заданных показателей с автоматическим выбором существенных признаков из полного набора признаков; с конкретно заданными признаками: функциями и аргументами; смешанным способом, при котором сначала строятся модели с заданным набором признаков, а затем они дополняются существенными признаками из оставшихся аргументов.

Необходимо заметить, что все три режима очень важны при проведении экономико-статистических исследований. Первый позволяет выделить из достаточно широкого набора признаков небольшое количество существенных для данного экономического показателя. Важность, необходимость решения такой задачи и применяемые методы выбора существенных признаков рассмотрены в гл. 2, § 4 настоящей книги. Метод автоматического выбора, используемый в данной программе, вытекает из алгоритма построения регрессии и будет рассмотрен позднее.

Второй режим позволяет проанализировать конкретные регрессионные модели, в которых набор аргументов выбран исходя из экономического анализа структуры изучаемого процесса, из необходимости учесть при построении динамических моделей те или другие лаговые переменные, фактор времени и пр.

Третий режим — комбинированный. В результате работы программы в этом режиме выбранный на этапе предварительного анализа «обязательный» набор признаков дополняется (а он, как правило, всегда нуждается в дополнении) другими аргументами, которые в сочетании с уже имеющимися дают наибольший эффект — они сильнее влияют на моделируемый показатель. Регрессионная модель, полученная при работе в этом режиме, позволяет сочетать качественные и количественные методы выбора существенных признаков.

3.12. Уравнение регрессии в данной программе строится не методом наименьших квадратов, как обычно, а с помощью сходящегося итерационного процесса построения уравнения регрессии по полному набору признаков методом последовательных приближений⁴.

Пусть имеется $m+1$ нормированных и центрированных (без нарушения общности) случайных величин y, x_1, x_2, \dots, x_m . Требуется построить линейное уравнение регрессии y на \vec{X} , где $\vec{X} = (x_1, x_2, \dots, x_m)$. Введем обозначения:

$$M(x_j x_i) = r_{ij}, M(y x_i) = r_i, i, j = 1, 2, \dots, m.$$

Предлагаемый алгоритм построения уравнения регрессии заключается в следующем.

1. Отыскивается i_0 такой, что

$$|r_{i_0}| = \max_{1 \leq i \leq m} |r_i|.$$

2. Строится парная регрессия y на x_{i_0} (для определенности при существовании двух разных индексов k и j , для которых $r_k = r_j = \max_i |r_i|$, выбираем наименьший).

3. Вычисляются парные коэффициенты корреляции r'_i между ξ_1 и $x_i, i = 1, 2, \dots, m$.

4. Отыскивается i_1 , такой что

$$|r'_{i_1}| = \max_{1 \leq i \leq m} |r'_i|.$$

5. Строится парная регрессия ξ_1 на x_{i_1} :

$$\xi_1 = \alpha_1 x_{i_1} + \xi_2,$$

где $\alpha = r_{i_1} / \sqrt{D(\xi_1)}, M(\xi_2) = 0, D(\xi_2) = 1 - \alpha_0^2 - \alpha_1^2$.

6. Пусть на n -м шаге найдена парная регрессия

$$\xi_n = \alpha_n x_{i_n} + \xi_{n+1},$$

где

$$\alpha_n = r_{i_n} / \sqrt{D(\xi_n)}, M(\xi_{n+1}) = 0,$$

$$D(\xi_{n+1}) = 1 - \sum_{j=0}^n \alpha_j^2.$$

⁴ Беккер А. В., Лукацкая М. Л. Выбор значимых факторов при моделировании экономических показателей.— В кн.: Вопросы построения и применения статистических моделей экономических показателей предприятий. Новосибирск, 1971, с. 79—88.

Тогда имеем:

$$\begin{aligned} y &= \alpha_0 x_{i_0} + \xi_1 \\ \xi_1 &= \alpha_1 x_{i_1} + \xi_2 \\ &\vdots \\ \xi_n &= \alpha_n x_{i_n} + \xi_{n+1} \end{aligned}$$

и, следовательно,

$$y = \sum_{j=0}^n \alpha_j x_{i_j} + \xi_{n+1}.$$

Сгруппировав члены с одинаковыми признаками (индексом i_j), на n -м шаге получим приближенные уравнения регрессии Y на X :

$$y = \sum_{i=1}^m \gamma_i x_i + \xi_{n+1}.$$

При этом некоторые значения γ_i могут быть равными нулю, т. е. не все признаки войдут в правую часть приближения. Точность этой аппроксимации можно оценить с помощью дисперсии остатка ξ_{n+1} , равной

$$D(\xi_{n+1}) = D\left(y - \sum_{j=0}^n \alpha_j x_{i_j}\right) = 1 - \alpha_0^2 - \alpha_1^2 - \dots - \alpha_n^2.$$

Как видно из алгоритма, эта величина остаточной дисперсии на каждом шаге уменьшается. Сходимость алгоритма доказана: линейная комбинация, получаемая методом последовательных приближений, в пределе совпадает с линейной комбинацией $\sum_{i=1}^m \beta_i x_i$, найденной методом наименьших квадратов.

3.13. Изложенный выше алгоритм построения последовательных приближений используется для определения набора существенных признаков моделируемого показателя. Если среди m анализируемых признаков имеется k существенно влияющих на изучаемый показатель, то по предлагаемому алгоритму именно эти значимые признаки будут включаться в отыскиваемые приближения к регрессии в первую очередь. Причем каждое такое добавление дает ощутимое уменьшение остаточной дисперсии, т. е. величина $\alpha_n^2 = D(\xi_n) - D(\xi_{n+1})$ будет такова, что исследова-

тель еще не сможет пренебречь информацией, которую несет включаемый признак. При дальнейшем применении алгоритма в уравнение начнут включаться и несущественно влияющие на показатель признаки, но для них изменения остаточной дисперсии будут незначительны (значения α_n^2 малы). Анализ последовательностей

$$\alpha_0^2, \alpha_1^2, \dots, \alpha_n^2,$$

$$i_0, i_1, \dots, i_n$$

позволяет определить количество значимых признаков и их перечень.

На каждом шаге итерационного процесса вычисляются и выдаются на печать максимальный коэффициент корреляции, коэффициенты последовательных приближений, отклонения и статистические характеристики отклонений действительных значений показателя от приближенных, а также значение разности между остаточными дисперсиями двух последовательных приближений. Кроме того, для проведения сравнительного анализа методом наименьших квадратов вычисляются коэффициенты соответствующих уравнений регрессии (обратная матрица вычисляется методом окаймления), выдаются на каждом шаге коэффициент множественной корреляции, коэффициент детерминации, отношение Фишера, средние квадратические отклонения регрессионных коэффициентов (для регрессий в натуральном и стандартизованном масштабах), а также отклонения теоретических значений от фактических, относительная ошибка регрессии и дисперсия отклонений.

Построение регрессии заканчивается на шаге, для которого разность остаточных дисперсий двух последовательных приближений не будет превышать заданного исследователем постоянного достаточно малого числа $\varepsilon > 0$.

АЛГОРИТМ ТРЕХШАГОВОГО МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ

3.14. Программа предназначена для решения задач на ЭВМ-220 и является вариантом программы, написанной А. И. Рогаткиным на языке «Альфа» для ЭВМ БЭСМ-6⁵.

⁵ Рогаткин А. И. Программа трехшагового метода наименьших квадратов.— В кн.: Вопросы построения и применения статистических моделей экономических показателей. Новосибирск, 1971, с. 123—124.

С ее помощью рассчитываются параметры взаимозависимых эконометрических уравнений двух- и трехшаговым методом наименьших квадратов (м. н. к.), который был предложен Зелнером и Тэйлом (Zelner, Theil, 1962).

3.15. Всякое сколько-нибудь сложное экономическое явление, как правило, описывается целой системой сложных взаимосвязей, отразить которые с помощью стандартной схемы регрессионного анализа не всегда оказывается возможным. Внутренний причинный механизм, структура связей при использовании одного уравнения регрессии остаются нераскрытыми. Комплексное представление об изучаемом явлении могут дать модели, включающие в себя систему статистических уравнений.

Вместо выделения одной зависимости и совокупности независимых переменных в системе уравнений устанавливается различие между совместно зависимыми переменными, влияние которых друг на друга должно быть исследовано, и предопределенными переменными, оказывающими влияние на совместно зависимые, однако, в свою очередь, не испытывающими влияния последних.

Совместно зависимые переменные, поскольку они определяются системой, называются эндогенными переменными. А в число предопределенных переменных помимо экзогенных (внешних для данной системы) входят запаздывающие эндогенные переменные.

3.16. Полная система должна содержать столько уравнений, сколько имеется эндогенных переменных. Число предопределенных переменных может быть каким угодно (больше или меньше числа эндогенных переменных). Таким образом, система уравнений с m эндогенными переменными ($y_{1t}, y_{2t}, \dots, y_{mt}$), k предопределенными ($x_{1t}, x_{2t}, \dots, x_{kt}$) и m возмущениями уравнений ($\eta_{1t}, \eta_{2t}, \dots, \eta_{mt}$) в общей форме может быть записана следующим образом:

$$y_{1t} = \beta_{10} + \beta_{12}y_{2t} + \dots + \beta_{1m}y_{mt} + \gamma_{11}x_{1t} + \dots + \gamma_{1k}x_{kt} + \eta_{1t}$$

$$\dots$$

$$y_{mt} = \beta_{m0} + \beta_{m1}y_{1t} + \dots + \beta_{m, m-1}y_{m-1, t} + \gamma_{m1}x_{1t} + \dots + \gamma_{mk}x_{kt} + \eta_{mt}$$

3.17. Использование классического метода наименьших квадратов для оценки параметров таких моделей приводит в общем случае к несостоятельным и асимптотически смещенным оценкам. Дело в том, что оказываются нарушенными условия, выполнение которых предполагает м. н. к.:

АНАЛИЗ ДИНАМИЧЕСКИХ РЯДОВ

а) отсутствие зависимости между переменными, отнесенными к числу аргументов; б) статистическая независимость переменных — аргументов от возмущений уравнений; в) отсутствие автокорреляции уравнений.

Для оценивания параметров систем структурных уравнений разработан целый ряд методов, в том числе двух- и трехшаговые методы наименьших квадратов, допускающие нарушение перечисленных выше требований. Двухшаговый м. н. к. предназначен для оценивания одного структурного уравнения в системе и использует ограниченную информацию, т. е. учитывает только ограничения, накладываемые моделью на оцениваемое уравнение. Трехшаговый м. н. к. учитывает информацию о ковариационной матрице структурных возмущений, а также ограничения, накладываемые моделью на остальные уравнения системы.

3.18 Программа, реализующая трехшаговый м. н. к., позволяет получать для каждого уравнения: 1) коэффициент Дарбина-Вотсона и множественный коэффициент корреляции; 2) оценки параметров уравнений и их средние квадратические отклонения; 3) разницу между эмпирическими и теоретическими значениями моделируемых показателей — как для каждого отдельного наблюдения, так и для всей совокупности наблюдений.

4.1. В приложении приводятся три программы, применяемые наиболее часто в анализе динамических рядов и построении динамических ЭСМ. Первая программа позволяет получать автокорреляционные и взаимокорреляционные функции системы случайных величин, заданных своими временными рядами; вторая — аппроксимирующую (сглаживающую) функцию динамического ряда в виде кривых определенного типа, т. е. определить тенденцию показателей; по третьей программе можно построить динамическую модель показателя, содержащую в качестве главных компонент тенденцию и авторегрессионную часть.

АВТОКОРРЕЛЯЦИОННЫЕ И ВЗАИМОКОРРЕЛЯЦИОННЫЕ ФУНКЦИИ

4.2. В качестве оценки нормированной взаимокорреляционной функции процессов $X(t)$ и $Y(t)$ принимается обычно

$$r(s) = \frac{n-s-1}{n-s} \frac{\sum_{i=1}^{n-s} (x_{i+s} - \bar{x}_s) (y_i - \bar{y}_s)}{\sqrt{\sum_{i=1}^{n-s} (x_{i+s} - \bar{x}_s)^2 \cdot \sum_{i=1}^{n-s} (y_i - \bar{y}_s)^2}},$$

где $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n$ — соответствующие реализации исследуемых процессов; \bar{x}_s, \bar{y}_s — средние арифметические этих рядов, сдвинутых на s наблюдений

$$\left(\bar{x}_s = \frac{1}{n-s} \sum_{i=1}^{n-s} x_{i+s}, \quad \bar{y}_s = \frac{1}{n-s} \sum_{i=1}^{n-s} y_i \right).$$

имствован из книги Дж. Юла и М. Кендалла (1960), а коэффициенты c для уравнений 12—14 находились по формулам, описанным Г. Крамером (1975).

Из 15 рассчитанных уравнений лучшим по формальному критерию — отражающим существующую в действительности зависимость между переменными y и x — считается то, для которого остаточная дисперсия минимальна. Она определяется по формуле

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n-1} - \left(\frac{\sum_{i=1}^n (y_i - \bar{y}_i)}{n} \right)^2,$$

где y_i — фактические значения переменной;
 \bar{y}_i — расчетные значения переменной;
 n — число наблюдений.

Программа рассчитывает и выдает на печать среднее значение функции (\bar{y}), его доверительный интервал, дисперсию функции (так называемую первоначальную дисперсию) и ее доверительный интервал. (Доверительные интервалы соответствуют определенному уровню значимости, заданному пользователем.) Для каждой из 15 кривых выдаются найденные параметры кривых (a, b, c, d, k), отклонения $y_i - \bar{y}_i$, среднее значение этих отклонений и остаточная дисперсия со своим доверительным интервалом, а также номер кривой, для которой остаточная дисперсия минимальна.

4.9. Программа «Подбор тенденций и прогноз» — вариант предыдущей. Как и в первом варианте, из 15 видов кривых выбирается та, которая дает наименьшую остаточную дисперсию. Первые 12 видов кривых обоих вариантов совпадают (как и их нумерация), остальные во втором варианте имеют следующий вид:

$$13. y = ab^x + c;$$

$$14. y = c(1 + bc^{-ax});$$

$$15. y = a \ln x + b.$$

В отличие от предыдущего варианта программы:

а) дополнительно вычисляются дисперсии коэффициентов уравнений $\sigma_a^2, \sigma_b^2, \sigma_c^2, \sigma_d^2$, и, следовательно, можно, рассчитав доверительные интервалы этих коэффициентов, решить вопрос о существенности включения того или иного слагаемого в модель;

б) подсчитываются прогнозные теоретические значения функций для заданных значений аргумента по любой кривой;

в) массивы исходной информации могут быть разной длины.

ПРОГРАММА «ДИНАМИКА» АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

4.10. В гл. 4 (§ 1) указывается как наиболее приемлемая и приемлемая в качестве динамической модели отдельного показателя авторегрессионная модель и дается схема ее построения: сначала находится тенденция временного ряда, а затем — авторегрессия для отклонений от тенденции. При построении подобных моделей можно использовать программу отыскания тенденции (описана в этом же приложении) и программу авторегрессионного анализа Г. П. Ульяновой (Статистические алгоритмы и программы, 1970, с. 252—264). Но в этом случае для окончательной модели вида

$$y(t) = \sum_{j=1}^s b_j y(t-j) + \sum_{i=0}^k a_i t^i + \varepsilon$$

параметры модели ($a_i, i=0, \dots, k; b_j, j=1, \dots, s$) не будут в определенном смысле оптимальными, так как найдены хотя и с помощью метода наименьших квадратов, но используемого дважды, последовательно.

Написана программа¹, позволяющая оценивать параметры модели сразу, применяя метод наименьших квадратов один раз. Но тогда вид модели — порядок авторегрессии и тип тенденции — необходимо задать заранее. По приводимой же в гл. 4 схеме порядок авторегрессии определяется в процессе исследования, в частности с помощью анализа автокорреляционной функции.

4.11. В качестве тенденции используются три типа кривых:

$$y = a_3 t^3 + a_2 t^2 + a_1 t + a_0;$$

$$y = \frac{a_3}{t^3} + \frac{a_2}{t^2} + \frac{a_1}{t} + a_0;$$

$$y = e^{a_1 t^2} + a_2 t^2 + a_1 t + a_0.$$

¹ Ульянова Г. П. Программа «Динамика» анализа временных рядов. — В кн.: Алгоритмы статистической обработки информации. Новосибирск, 1974, с. 40—55.

Алгоритмом предусмотрена возможность фиксировать один или группу коэффициентов из множества $\{a_3, a_2, a_1, a_0\}$ на нулевом уровне и тем самым получать различные типы тенденций. (Остальные коэффициенты оцениваются в процессе построения авторегрессионной модели.)

4.12. Авторегрессионная часть модели имеет вид $\sum_{j=1}^s b_j y(t-j)$. Некоторые из этих коэффициентов тоже могут быть заданы равными нулю, т. е. можно исследовать и оценивать неполные схемы авторегрессии.

4.13. Таким образом, имеем следующие три типа авторегрессионных моделей, которые можно построить, используя данную программу:

$$y(t) = a_3 t^3 + a_2 t^2 + a_1 t + a_0 + b_1 y(t-1) + b_2 y(t-2) + \dots + b_s y(t-s),$$

$$y(t) = a_3/t^3 + a_2/t^2 + a_1/t + a_0 + b_1 y(t-1) + b_2 y(t-2) + \dots + b_s y(t-s).$$

$$y(t) = e^{a_3 t^3 + a_2 t^2 + a_1 t + a_0} \cdot y_{(t-1)}^{b_1} \cdot y_{(t-2)}^{b_2} \cdot \dots \cdot y_{(t-s)}^{b_s}.$$

В последней модели авторегрессионная часть имеет специфическую форму $\left(\prod_{j=1}^s y^{b_j}(t-j) \right)$, соответствующую тенденции вида $e^{a_3 t^3 + a_2 t^2 + a_1 t + a_0}$, так как метод наименьших квадратов применяется не к исходному временному ряду, а к прологарифмированному, образованному логарифмами членов исходного ряда.

Программа позволяет получить в чистом виде и тенденцию, и авторегрессионную схему любого порядка — для этого достаточно задать равными нулю все коэффициенты b_j ($j=1, \dots, S$) или ($i=0, 1, 2, 3$).

4.14. На печать выдаются коэффициенты авторегрессионной модели ($b_j, j=1, \dots, S, a_i, i=0, \dots, 3$), теоретические значения функции y , их отклонения от фактических, среднее значение, дисперсия и среднеквадратическая ошибка отклонений.

Использованные в программе стандартные процедуры ЦИФЛ, БУКЛ, ТЕКСТ позволяют организовать печать результатов в виде таблиц, употребляемых в отчетах. Количество экземпляров листов с таблицами задается составителем задачи. Ему же предоставляется возможность выбора количества значащих цифр при печати результатов.

ЛИТЕРАТУРА

Аганбегян А. Г. О внедрении экономико-математических методов и электронно-вычислительных машин в разработку перспективных планов развития и размещения производства в отдельных отраслях промышленности. — В кн.: Оптимальное отраслевое планирование в промышленности. Новосибирск, 1970, с. 3—13.

Аганбегян А. Г., Бородкин Ф. М. О применении математической статистики в экономическом анализе и планировании. — В кн.: Научные труды Института экономики и организации промышленного производства СО АН СССР и НГУ, вып. VII. Новосибирск, 1963, с. 6—23.

Аганбегян А. Г., Багриновский К. А., Гранберг А. Г. Система моделей народнохозяйственного планирования. М., 1972. 351 с.

Аганбегян А. Г., Майер В. Ф. Заработная плата в СССР, М., 1959. 240 с.

Айвазян С. А. Статистическое исследование зависимостей. М., 1968. 228 с.

Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. М., 1974. 240 с.

Айзерман М. А., Браверман Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. М., 1970. 384 с.

Алгоритмы статистической обработки информации. Под ред. М. Л. Лукацкой. Новосибирск, 1974. 142 с.

Андерсон Т. Введение в многомерный статистический анализ. М., 1963. 500 с.

Аркадьев А. Г., Браверман Э. М. Обучение машины классификации объектов. М., 1971. 192 с.

Базилевич А. А. Постановка задачи линейного программирования с применением данных корреляционного анализа. — «Экономика и математические методы», 1967, т. 3, вып. 1, с. 83—87.

Барабаш Ю. Л., Варский Б. В., Кириченко В. С., Зиновьев В. Т., Сапегин В. Ф. Вопросы статистической теории распознавания. М., 1967. 399 с.

Берж К. Теория графов и ее применение. М., 1962. 319 с.

Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., 1965. 476 с.

- Бонгард М. М.** Проблема узнавания. М., 1967. 319 с.
- Бородкин Ф. М.** Статистическая оценка связи экономических показателей. М., 1968. 204 с.
- Бородюк В. П.** Проверка однородности статистических данных в регрессионном анализе.— В кн.: Проблемы планирования эксперимента. М., 1967, с. 7—12.
- Бородюк В. П., Голяс Ю. Е.** Линейное программирование для оптимизации статистической модели.— В кн.: Доклады научно-технической конференции по итогам НИР за 1968—1969 гг. Секция автоматической, вычислительной и измерительной техники, подсекция автоматического управления. М., 1970, с. 73—83.
- Браверман Э. М.** Метод потенциальных функций в задаче обучения машины распознаванию образов без учителя.— «Автоматика и телемеханика», 1966, № 10, с. 100—121.
- Браверман Э. М.** Методы экстремальной группировки и задача выделения существенных факторов.— «Автоматика и телемеханика», 1970, № 1, с. 123—133.
- Браверман Э. М., Дорофеев А. А., Лумельский В. Н., Мучник И. В.** Диагонализация матрицы связи и выделение скрытых факторов.— В кн.: Проблемы расширения возможностей автоматов, вып. 1. М., 1971, с. 42—79.
- Браун М.** Теория и измерение технического прогресса. М., 1971. 208 с.
- Браунли К. А.** Статистические исследования в производстве. М., 1949. 228 с.
- Будницкий И. М.** Экономика угольной промышленности. М., 1959. 236 с.
- Вайншток М. Д.** Оптимизация многосвязных объектов управления с перекрестными связями при наличии помех как задача стохастического программирования.— В кн.: Управление многосвязными системами. 3-е Всесоюз. совещание (Москва, апрель, 1973), ч. 1. М., 1974, с. 22—25 (Рефераты докладов).
- Ван дер Варден В. Л.** Математическая статистика. М., 1960. 434 с.
- Венецкий И. Г., Кильдишев Г. С.** Основы теории вероятностей и математической статистики. Изд. 3-е. М., 1968. 360 с.
- Вентцель Е. С.** Теория вероятностей. М., 1969. 576 с.
- Верховская Л. А.** Однородность средних значений и корреляционных матриц многомерных выборок.— «Изв. АН СССР. Техническая кибернетика», 1971, № 5, с. 171—176.
- Вишнев С. М.** Экономические параметры (введение в теорию показателей экономических систем и моделей). М., 1968. 189 с.
- Волков О. И.** Показатели технического уровня производства.— «Вопросы экономики», 1970, № 8, с. 58—67.
- Вопросы построения и применения статистических моделей экономических показателей предприятий.** Под ред. Б. Б. Розина, М. Л. Лукацкой. Новосибирск, 1971, ч. I, 190 с.; ч. II, 290 с.
- Вопросы экономико-статистического моделирования и прогнозирования в промышленности.** Под ред. Б. Б. Розина. Новосибирск, 1970. 290 с.
- Данциг Дж., Вулф Ф.** Алгоритмы разложения для задачи линейного программирования.— «Математика», 1964, т. 8, № 1, с. 151—160.
- Дворецкий Н. М., Фрайман Я. В., Трушина Г. С., Бедин В. И.** Прогнозирование роста производительности труда для Ленинско-Зеловской группы шахт Кузбасса.— «Экономика угольной промышленности», 1973, № 5, с. 23—26.
- Дорофеев А. А.** Алгоритмы автоматической классификации (обзор литературы).— В кн.: Проблемы расширения возможностей автоматов, вып. 1. М., 1971, с. 5—41.
- Дрейпер И., Смит Г.** Прикладной регрессионный анализ. М., 1973. 392 с.
- Дружинин Н. К.** Математическая статистика в экономике. М., 1974. 264 с.
- Езекиел М., Фокс К.** Методы анализа корреляций и регрессий (линейных и нелинейных). М., 1966. 559 с.
- Елкина В. Н.** Программа «Крб» для ЭВМ БЭСМ-6.— В кн.: Вычислительные системы, вып. 45. Новосибирск, 1971, с. 47—55.
- Елкина В. Н., Загоруйко Н. Г.** Об алфавите объектов распознавания.— В кн.: Вычислительные системы, вып. 22. Новосибирск, 1966, с. 59—77.
- Елкина В. Н., Загоруйко Н. Г.** Количественные критерии качества таксономии и их использование в процессе принятия решений.— В кн.: Вычислительные системы, вып. 36. Новосибирск, 1969, с. 29—46.
- Житков Г. Н.** Некоторые методы автоматической классификации (обзор).— В кн.: Структурные методы опознавания и автоматическое чтение. М., 1970, с. 68—85.
- Загоруйко Н. Г.** Классификация задач распознавания образов.— В кн.: Вычислительные системы, вып. 22. Новосибирск, 1966, с. 3—20.
- Загоруйко Н. Г.** Общие свойства задач распознавания образов.— В кн.: Вычислительные системы, вып. 44. Новосибирск, 1971, с. 3—11.
- Загоруйко Н. Г.** Методы распознавания и их применение. М., 1972. 206 с.
- Зайцев Б. Ф., Чирков В. Г.** Техничко-экономический уровень производства. М., 1972. 190 с.
- Касавин А. Д.** Методы кусочной аппроксимации в задаче идентификации сложных объектов. Автореф. дис. М., 1974. 22 с.
- Кендалл М. Дж., Стьюарт А.** Теория распределений. М., 1966. 588 с.
- Ковалевский В. А.** Распознавание образов: эвристика или наука? (Обзор). Киев, 1970. 94 с.
- Козлов Л. А.** Оптимальное планирование развития и размещения отраслей промышленности. Новосибирск, 1970. 176 с.
- Корнаи И., Липтак Т.** Планирование на двух уровнях.— В кн.: Применение математики в экономических исследованиях, т. 3. М., с. 107—136.
- Крамер Г.** Математические методы статистики. М., 1975. 648 с.
- Кулаков А. К., Алексеев А. М.** Мобильные (плавающие) беспромхозы и эффективность их применения. М., 1971. 60 с.
- Кулик В. Т.** Цифровое моделирование сложных систем. Киев, 1964. 199 с.
- Кулик В. Т.** Алгоритмизация объектов управления. Киев, 1968. 363 с.
- Кульбак С.** Теория информации и статистики. М., 1967. 408 с.

Лбов Г. С. Выбор эффективной системы зависимости признаков.— В кн.: Вычислительные системы, вып. 19. Новосибирск, 1965, с. 21—30.

Ли Р. Оптимальные оценки, определение характеристик и управление. М., 1966. 176 с.

Лизер С. Эконометрические методы и задачи. М., 1971. 141 с.

Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е. М., 1958. 333 с.

Литвинов Б. М. Об одном итерационном методе в задаче аппроксимации функций по конечному числу наблюдений.— «Автоматика и телемеханика», 1966, № 4, с. 104—113.

Лоули Д., Максвелл А. Факторный анализ как статистический метод. М., 1967. 144 с.

Лукомский Я. П. Теория корреляции и ее применение к анализу производства. М., 1961. 375 с.

Лумельский В. Я. Группировка параметров на основе квадратной матрицы.— «Автоматика и телемеханика», 1970, № 1, с. 133—143.

Маршак М. И. Адаптация корреляционных моделей технологического процесса.— В кн.: Проблемы теории статистики. М., 1970, с. 94—113 (Труды МЭСИ).

Матусhevский В. В. Определение необходимого объема эксперимента в задаче регрессионного анализа, Томск, 1965. 47 с.

Методические положения оптимального отраслевого планирования в промышленности. Новосибирск, 1972. 312 с.

Миллс Ф. Статистические методы. М., 1958. 798 с.

Милиц А. А. Экономическая оценка естественных ресурсов. М., 1972. 303 с.

Миркин Б. Г. Об одном подходе к обработке нечисловых данных.— В кн.: Математические методы моделирования и решения экономических задач. Новосибирск, 1969, с. 140—150.

Миркин Б. Г. Об одном подходе к анализу первичной социологической информации (Доклад на VII Международном социологическом конгрессе. Варна, сентябрь 1970 г.). Новосибирск, 1970. 5 с.

Миркин Б. Г., Черный Л. Б. Аксиоматический подход к измерению близости между различными классификациями объектов данного множества.— В кн.: Математические методы моделирования экономических задач. Новосибирск, 1969, с. 151—159.

Миркин Б. Г., Черный Л. Б. Некоторые свойства пространства упорядоченных и неупорядоченных разбиений.— В кн.: Математический анализ народнохозяйственных моделей, ч. III. Новосибирск, 1972, с. 126—147.

Митропольский А. К. Техника статистических вычислений. М., 1971. 576 с.

Моргенштерн О. О точности экономико-статистических наблюдений. М., 1968. 293 с.

Мот Ж. Статистические предвидения и решения на предприятии. М., 1966. 501 с.

Мюллер Д., Шусслер К. Статистические методы в социологии, ч. II. М., 1968. 183 с.

Надь Дж. Распознавание образов (обзор).— ТИИЭР, 1968, т. 56, № 5, с. 57—86.

Немчинов В. С. Экономико-математические методы и модели. М., 1962. 409 с.

Новожилов В. В. Проблема измерения затрат и результатов в оптимальном планировании. М., 1972. 434 с.

Проблемы экономико-статистического анализа и моделирования промышленного производства. Под ред. Б. Б. Розина. Новосибирск, 1969. 280 с.

Райбман Н. С., Чадеев В. М. Адаптивные модели в системах управления. М., 1966. 159 с.

Райхман У. Дж. Применение статистики. М., 1969. 296 с.

Рао С. Р. Линейные статистические методы и их применение. М., 1968. 547 с.

Распознавание образов в социальных исследованиях. Под ред. Н. Г. Загоруйко, Т. И. Заславской. Новосибирск, 1968. 195 с.

Распознавание образов и регрессионный анализ в экономических исследованиях. Под ред. Б. Б. Розина, Н. М. Журавель. Новосибирск, 1972. 280 с.

Распознавание образов в экономико-статистическом моделировании. Под ред. Б. Б. Розина, М. Л. Лукацкой. Новосибирск, 1974. 225 с.

Реусс Г. Анализ производительности (экономические основы и статистическая методика). М., 1963. 251 с.

Розин Б. Б. Теория распознавания образов в экономических исследованиях. М., 1973. 224 с.

Розин Б. Б. Статистическое моделирование экономических показателей. Новосибирск, 1976. 136 с.

Розин Б. Б., Гейфман Р. С. Экономико-математические исследования на металлургическом заводе. М., 1966. 264 с.

Розин Б. Б., Лукацкая М. Л., Овчинникова Н. В. Дискретно-непрерывные модели экономических показателей лесозаготовительного производства.— В кн.: Оптимизационные и экономико-статистические методы в перспективном отраслевом планировании. Новосибирск, 1975, с. 163—170.

Себестяна Г. С. Процессы принятия решений при распознавании образов. Киев, 1965. 151 с.

Смирнов Н. В., Дуини-Барковский М. В. Краткий курс математической статистики для технических приложений. М., 1959. 436 с.

Снедекор Дж. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. М., 1961. 503 с.

Статистические методы в экономическом анализе производства. Под ред. Б. Б. Розина. Новосибирск, 1968. 199 с.

Статистические алгоритмы и программы. Под ред. Ф. М. Бородкина, М. Л. Лукацкой. Новосибирск, 1970. 290 с.

Статистическое моделирование и прогнозирование технико-экономических показателей предприятий. Новосибирск, 1974. 192 с.

Суслев И. П. Методология экономических исследований. М., 1974. 334 с.

Тинтнер Г. Введение в эконометрику. М., 1965. 361 с.

Турбович И. Т., Гитис В. Г., Маслов В. К. Опознавание образов (детерминированно-статистический подход). М., 1971. 246 с.

Уилкс С. Математическая статистика. М., 1967. 632 с.

Устюжанинов В. Л. Информационные меры и их использование в социологическом анализе.— В кн.: Измерение и моделирование в социологии. Новосибирск, 1969, с. 62—86.

При $s=0$ эта формула дает парный коэффициент корреляции между рядами $x_i (i=1, \dots, n)$ и $y_i (i=1, \dots, n)$, при $s \neq 0$ — парный коэффициент корреляции между сдвинутыми рядами; при замене ряда y_i на x_i получается парный коэффициент корреляции ряда x_i и того же ряда, только сдвинутого на s шагов, т. е. коэффициент автокорреляции s -го порядка.

4.3. Известна программа (Статистические алгоритмы и программы, 1970, с. 235—251), которая по совокупности временных рядов подсчитывает для каждого из них среднее значение и дисперсию, а также их доверительные интервалы; значения автокорреляционных функций для каждого ряда и для сдвигов от 1-го до определенного, задаваемого исследователем порядка; значения взаимокорреляционных функций для всех пар заданных рядов и тех же сдвигов; доверительные интервалы с определенным уровнем значимости для всех коэффициентов — автокорреляционных и взаимокорреляционных.

4.4. Анализ автокорреляционных и взаимокорреляционных функций позволяет определить характер авторегрессионной схемы и лаговые переменные при построении динамической экономико-статистической модели как отдельного объекта, так и совокупности объектов (гл. 4).

ОПРЕДЕЛЕНИЕ ТЕНДЕНЦИИ (ТРЕНДА) ДЛЯ ДИНАМИЧЕСКИХ РЯДОВ

4.5. Как уже указывалось (гл. 4), большое значение при экономико-статистическом моделировании имеет анализ динамических рядов экономических показателей и построение частных динамических моделей. В программах «Подбор тенденций для динамических рядов (в кодах ЭВМ М-20)»; «Подбор тенденций и прогноз (на α -языке)» (Статистические алгоритмы и программы. Новосибирск, 1970, с. 194—234) реализован один из распространенных методов анализа рядов динамики — метод наименьших квадратов, позволяющий отыскать сглаживающую функцию в явном виде и представить изменение показателя во времени как сумму некоторой детерминированной функции аргумента t и случайного компонента ($y(t) = f(t) + \varepsilon$).

Динамические модели подобного типа могут использоваться для прогнозирования агрегированных, интегральных показателей более крупной системы, чем отдельное промышленное предприятие (например, для оценки спроса на продукцию отрасли — в абсолютном или относительном выражении).

4.6. При анализе построенных ЭСМ важно исследовать отклонения действительных значений показателей от теоретических, т. е. значения $\{e_i\}$, и отыскать функции распределения случайной величины ε . Если последовательность $\{e_i\}$, $i=1, \dots, T$ можно считать выборкой какой-то случайной величины, то, значит, нам удалось отыскать простое преобразование, которое приводит к процессу, близкому к последовательности независимых случайных величин, и может быть использовано не только для анализа показателя в исследуемый период, но при определенных условиях и для прогноза.

4.7. Упомянутые выше программы могут быть использованы несколько шире, чем отыскание тенденций при анализе временных рядов. В качестве аргумента в них может выступать любая переменная, а не только t (обычно принимает значения натурального ряда), поэтому можно получать параметры и оценки парных регрессий для любых двух показателей, т. е. построить модель типа $y = f(x) + \varepsilon$, которая может быть использована при проведении корреляционно-регрессионного анализа для изучения и выбора формы парных связей (гл. 2, § 4, гл. 4, § 2, 4).

4.8. Для отыскания зависимости $y = f(x)$ в программе «Подбор тенденций для динамических рядов» предусмотрено 15 видов кривых:

- | | |
|------------------------------------|---|
| 1. $y = ax + b$; | 9. $y = a/x^3 + b/x^2 + c/x + d$; |
| 2. $y = a/x + b$; | 10. $y = a \ln^3 x + b \ln^2 x + c \ln x + d$; |
| 3. $y = 1/(a/x + b)$; | 11. $y = 1/(ax^3 + bx^2 + cx + d)$; |
| 4. $y = ax^2 + bx + c$; | 12. $y = ax^b + c$; |
| 5. $y = a/x^2 + b/x + c$; | 13. $y = ae^{bx} + c$; |
| 6. $y = a \ln^2 x + b \ln x + c$; | 14. $y = ab^x + c$; |
| 7. $y = 1/(ax^2 + bx + c)$; | 15. $y = k/(1 + be^{-ax})$. |
| 8. $y = ax^3 + bx^2 + cx + d$; | |

Параметры уравнений a, b, c, d, k определяются: для уравнений с 1-го по 11-е методом наименьших квадратов, для логистической кривой (уравнение 15) расчет их за-

ОГЛАВЛЕНИЕ

Введение	3
Раздел I. Методологические и методические вопросы экономико-статистического моделирования	14
<i>Глава 1.</i> Основные положения	—
§ 1. Объект и предмет исследования	—
§ 2. Задача и основные этапы экономико-статистического моделирования	16
§ 3. Классификация экономико-статистических моделей	21
§ 4. Основные направления применения ЭСМ	26
§ 5. Методы экономико-статистического исследования	30
<i>Глава 2.</i> Постановка задачи экономико-статистического моделирования. Предварительный анализ исходной информации	34
§ 1. Формулировка целей исследования. Априорный анализ процесса функционирования объекта	35
§ 2. Организация экспертного опроса и анализ его результатов	39
§ 3. Сбор и подготовка исходной информации	48
§ 4. Предварительный статистический анализ и преобразование исходной информации	53
<i>Глава 3.</i> Типы статистических моделей и методы их построения	60
§ 1. Выбор рационального типа модели	—
§ 2. Исследование совокупности на однородность. Построение регрессионной модели	63
§ 3. Классификация совокупности	68
§ 4. Построение дискретной модели	77
§ 5. Построение дискретно-непрерывной модели	84
<i>Глава 4.</i> Некоторые вопросы построения динамических ЭСМ	92
§ 1. Динамические модели отдельных объектов (частные динамические ЭСМ)	94
§ 2. Построение динамических моделей совокупности объектов (на базе частных статических ЭСМ)	108
§ 3. Построение общей динамической модели совокупности объектов (на базе частных динамических ЭСМ)	120

§ 4. Построение общей динамической модели по исходной информации малой размерности	122
Раздел II. Вопросы применения экономико-статистических моделей в анализе, нормировании и планировании промышленного производства	128
<i>Глава 5.</i> Применение ЭСМ в техническом нормировании, внутривзаводском планировании и управлении	—
§ 1. Статистическая оценка внутривзаводских нормативов	129
§ 2. Выбор экономически выгодных режимов управления агрегатом	136
<i>Глава 6.</i> Моделирование основных этапов развития промышленного предприятия	149
§ 1. Оценка уровня и характера распределения капитальных вложений в строительство промышленных объектов	150
§ 2. Моделирование и прогнозирование процесса освоения производственных мощностей	154
§ 3. Модель функционирования предприятия	159
<i>Глава 7.</i> Применение экономико-статистических моделей в межзаводском анализе и планировании	166
§ 1. Сравнительный технико-экономический анализ на основе ЭСМ	—
§ 2. Статистическое соизмерение объективных условий производства	169
§ 3. Статистическая оценка технического уровня производства	184
§ 4. Пример использования ЭСМ в оптимальном отраслевом планировании	189
<i>Приложение 1.</i> Исследование выборочной совокупности объектов на однородность	198
<i>Приложение 2.</i> Исследование качества классификации	210
<i>Приложение 3.</i> Алгоритмы корреляционно-регрессионного анализа	216
<i>Приложение 4.</i> Анализ динамических рядов	225
<i>Литература</i>	231

**ЭКОНОМИКО-СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
В ПРОМЫШЛЕННОСТИ**

(методологические и методические вопросы)

Ответственный редактор
Бенциан Борисович Розин

Редактор *И. Г. Зыкова*
Художник *В. И. Пинаев*
Технический редактор *Г. Я. Герасимчук*
Корректоры *С. В. Блинова, М. В. Ржевцева*

Сдано в набор 23 апреля 1976 г. Подписано в печать 24 декабря 1976 г.
МН 02109. Формат 84×108¹/₃₂. Бумага типографская № 3. 7,5 печ. л., 12,6
усл.-печ. л., 12,9 уч.-изд. л. Тираж. 2900 экз. Заказ № 104. Цена 1 р. 13 к.

Издательство «Наука», Сибирское отделение. 630099, Новосибирск, 99, Со-
ветская, 18.

4-я типография издательства «Наука». 630077. Новосибирск. 77, Станислав-
ского, 25.