

УДК 504.003; 504.001.89
КП 24741741
№ Госрегистрации 0197U005424
Инв.№

Национальная Академия наук Украины
Министерство образования Украины
Международный научно-учебный центр ЮНЕСКО/МПИ
информационных технологий и систем
03680, ГСП, Киев-187, проспект Глушкова, 40
тел.(044) 266-25-49

УТВЕРЖДАЮ
Директор МНУЦИТС
чл.-корр. АТН Украины
_____ Гриценко В.И.
1999.12.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
РАЗРАБОТАТЬ МЕТОДЫ, АЛГОРИТМЫ И СООТВЕТСТВУЮЩЕЕ
ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ МОДЕЛИРОВАНИЯ,
ПРОГНОЗИРОВАНИЯ И КООРДИНИРУЮЩЕГО УПРАВЛЕНИЯ
РЕГИОНАЛЬНЫМИ ЭКОЛОГО-ЭКОНОМИЧЕСКИМИ СИСТЕМАМИ
(заключительный)
Часть 1

Руководители НИР
зав.отделом
докт.техн.наук
Советник дирекции
член-корр.НАНУ

В.С.Степашко
А.Г.Ивахненко

1999

Рукопись закончена
Результаты этой работы рассмотрены Ученым Советом
МНУЦИТС, протокол от _____ №

СПИСОК АВТОРОВ

Зав.отделом
докт.техн.наук

В.С.Степашко
(реферат, предисловие,
введение, часть 1,
заключение)

Руководитель НИР
советник дирекции
член-корр. НАНУ

А.Г.Ивахненко
(раздел 1, подраздел
3.5 части 1)

Старший науч.сотр.
канд.техн.наук

Ю.В.Коппа
(раздел 3 части 1)

Научный сотрудник

Ю.В.Костенко
(подраздел 1.5
части 1)

РЕФЕРАТ

Объект исследования - эколого-экономические системы как многофакторные системы сложных взаимосвязанных процессов.

Цель работы - разработать методы, алгоритмы и программное обеспечение для моделирования, прогнозирования и сбалансированного управления сложными эколого-экономическими системами.

Методы исследования - структурная идентификация прогнозирующих моделей на основе метода группового учета аргументов (МГУА), теория анализа и синтеза многосвязных балансных систем управления и аппарат многоуровневых интегральных моделей развивающихся систем.

Исследованы закономерности уменьшения сложности (числа оцениваемых параметров) оптимальной модели при увеличении уровня (дисперсии) шума; разработана методика оценки эффективности применения практических критериев качества структур в задаче минимизации дисперсии ошибки прогнозирования; установлена адекватность применения критериев МГУА, основанных на разбиении выборки, для структурной идентификации помехоустойчивых прогнозирующих моделей.

Разработан новый комбинаторно-селекционный алгоритм МГУА с полиномиальной зависимостью объема вычислений от числа переменных. Предложена методология разработки программного обеспечения, функциональной архитектуры и структуры диалога программных систем моделирования по экспериментальным данным на основе МГУА. Приведены результаты решения ряда практических задач моделирования и прогнозирования макроэкономических процессов в Украине.

Исследован механизм саморегулирования и взаимодействия экологических и экономических подсистем единой замкнутой системы. Подобные системы преобразуют материальные потоки и являются многосвязными динамическими системами балансного типа. Они описываются вырожденными дифференциальными уравнениями, для их исследования предложены новые методы выявления условий возникновения системных эффектов: структурно-динамических резонансов, динамических дефектов (уменьшения порядка характеристического полинома) отрицательного перерегулирования и др. Найдены условия компенсации неконтролируемых возмущений (инвариантности) и координации переменных (регулирования соотношений) таких систем.

Исследована проблема описания рационального, экологически сбалансированного функционирования и развития экономики региона на основе применения интегральных моделей динамических систем с управляемой памятью. Детально рассмотрена макроэкономическая модель В.М.Глушкова, приведена экономическая интерпретация функций, входящих в эту модель, и ее обобщения. Некоторые из актуальных задач исследованы в данном отчете: многокритериальная оптимизация управления двухуровневой эколого-экономической системой "Центр-Предприятие"; оптимизация обновления производства (замена его на более эффективное и экологически безопасное) при заданном экологическом балансе.

Прогнозные предложения о развитии области исследования - теория и применение МГУА как метода оперативного прогнозирования; исследование системных эффектов в линейных балансных системах; изучение развивающихся систем на основе аппарата интегральных уравнений.

ЭКОЛОГИЯ, ЭКОНОМИКА, МОДЕЛИРОВАНИЕ, МГУА, БАЛАНСНАЯ СИСТЕМА, КООРДИНАЦИЯ, УПРАВЛЕНИЕ, ИНТЕГРАЛЬНОЕ УРАВНЕНИЕ, ПРОГРАММНАЯ СИСТЕМА.

СОДЕРЖАНИЕ

Предисловие	
Часть 1	
РАЗРАБОТКА АЛГОРИТМОВ И ПРОГРАММНЫХ СРЕДСТВ МОДЕЛИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ МЕТОДА ГРУППОВОГО УЧЕТА АРГУМЕНТОВ	
Введение	
1 Задача моделирования и прогнозирования экономических процессов по статистическим данным	
1.1 Моделирование экономических процессов по данным наблюдений в условиях неопределенности как задача структурной идентификации	
1.2 Сравнительная характеристика существующих подходов к построению моделей по данным наблюдений	
1.3 МГУА как оригинальный метод автоматизированного построения моделей	
1.3.1 Классы моделей, применяемые в МГУА	
1.3.2 Сравнительный анализ генераторов структур моделей.....	
1.3.3 Внешние критерии качества моделей.....	
1.4 Качественное сравнение МГУА и других статистических методов	
1.5 Многоэтапный (комбинаторно-селекционный) алгоритм с полиномиальной сложностью перебора моделей.....	
2 Исследование закономерностей выбора структуры модели по условию минимума дисперсии ошибки	
2.1 Постановка задачи структурной идентификации моделей с минимальной дисперсией ошибки.....	
2.3. Закономерности выбора оптимальных структур моделей в зависимости от уровня шума.....	
2.3.1. Случай усреднения по всем возможным реализациям шума.....	
2.3.2. Случай отдельной реализации шума	
2.4. МГУА как эффективный метод построения моделей с минимальной дисперсией ошибки.....	
2.5 Анализ проблемы разбиения выборки для алгоритмов МГУА	
3 Моделирование и прогнозирование макроэкономических процессов на основе МГУА.....	
3.1 Методика выбора наиболее существенных аргументов.....	
3.2 Пример 1: поступления в государственный бюджет Украины.....	
3.3 Пример 2: объем производства легкой промышленности.....	
3.4 Пример 3: процесс изменения индекса инфляции	
3.5 Индуктивный компьютерный советчик для прогнозирования макроэкономики Украины.....	
3.6 Сравнение прогнозирующих свойств моделей, полученных с применением регрессионного анализа и МГУА	

4 Разработка основ методологии создания интерактивных систем моделирования по данным наблюдений.....	
4.1 Методика структуризации знаний в области моделирования по экспериментальным данным	
4.2 Конструирование обобщенных многофункциональных вычислительных модулей	
4.3 Проектирование функциональной и диалоговой структуры программной системы моделирования.....	
4.4 Общая характеристика возможностей программной системы АСТРИД	
Выводы	
Перечень ссылок.....	

ЧАСТЬ 1

РАЗРАБОТКА АЛГОРИТМОВ И ПРОГРАММНЫХ СРЕДСТВ МОДЕЛИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ МЕТОДА ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

ВВЕДЕНИЕ

В первой части, состоящей из четырех разделов, анализируется задача построения прогнозирующих моделей сложных процессов по статистическим данным, создаются основы теории структурной идентификации моделей с минимальной дисперсией ошибки и анализируется в рамках этой теории эффективность индуктивного подхода, реализованного в виде известного метода группового учета аргументов (МГУА). Разрабатываются и применяются на реальных примерах алгоритмы и программные средства моделирования экономических процессов на основе МГУА.

В первой части отчета исследуются задачи обработки статистических данных наблюдений за процессами различной природы - экологическими, экономическими, технологическими - с целью построения моделей для описания этих процессов, прогнозирования, принятия решений по управлению и т.д. При этом решаются задачи построения моделей с оцениванием как структуры, так и параметров (структурной идентификации) в условиях короткой выборки зашумленных наблюдений и неопределенности состава факторов, существенно влияющих на целевой показатель.

Структурная идентификация рассматривается здесь как задача поиска структуры модели с минимальной дисперсией ошибки прогнозирования. Отметим, что задачи моделирования и прогнозирования тесно связаны между собой, но не тождественны. Задачей *моделирования* является построение математических моделей для количественного описания связи между целевыми (выходными) показателями (зависимыми переменными) моделируемых процессов и входными (независимыми) переменными (факторами). *Прогнозирование* – это количественное предсказание будущего развития процесса по моделям, полученным тем или иным методом. При этом задача прогнозирования может выступать как экстраполяция закономерного развития моделируемых процессов на будущий промежуток времени или как *имитационное моделирование*, то есть применение математических моделей для исследования изменения моделируемых процессов с учетом разных вариантов поведения независимых факторов (нормативное прогнозирование).

Традиционно наиболее распространенными являются две группы методов, на основе которых строят модели для прогнозирования. Экономико-математическое моделирование – это описание взаимодействия социально-экономических процессов разного уровня на основе наиболее полного изучения внутренних механизмов функционирования моделируемых процессов и представления полученных знаний в форме схем прямого расчета материально-финансовых потоков и определенных балансных соотношений. Эконометрическое моделирование – это получение моделей с помощью оценивания

параметров по выборкам наблюдений с применением регрессионного анализа при структурах моделей (регрессионных зависимостях), заданных с точностью до параметров.

В целом экономико-математический и эконометрический подходы отличаются, поскольку последний ориентирован на включение в модель лишь самых существенных в конкретных условиях факторов, а не всех, которые могут влиять на целевой показатель. Однако обе эти группы методов следует отнести к дедуктивным, методология которых состоит в переходе “от общего к частному”, то есть от изучения общих закономерностей к построению моделей конкретных явлений и процессов.

В отличие от этого, *индуктивное* моделирование базируется на принципе “от частного к общему”, т. е. от эмпирических данных о конкретных процессах до построения моделей с помощью информационных технологий “извлечения” знаний из данных. При этом математические модели строятся по выборкам наблюдений с помощью автоматического генерирования и сравнения большого количества различных моделей с выбором наилучшей в смысле соответствия той информации, которая неявно содержится в данных. Цель такого подхода – выявление скрытых закономерностей, то есть поиск описания закономерных причинно-следственных связей между элементами исследуемых процессов на основе статистических данных, в которых эти связи объективно отражены.

Как уже было сказано, при использовании эконометрических методов существенной проблемой является задание структуры модели. Эта проблема несущественна для индуктивного подхода к моделированию, который реализован в методе группового учета аргументов (МГУА), который фактически является методом поиска закономерностей с автоматическим выбором структуры и параметров модели на основе информации короткой выборки данных. В его основу положены принципы внешнего дополнения, автоматической генерации и последовательной селекции усложняемых структур моделей.

В этой части отчета формулируется задача построения прогнозирующих моделей сложных процессов по статистическим данным, создаются основы теории структурной идентификации моделей с минимальной дисперсией ошибки и анализируется в рамках этой теории эффективность индуктивного подхода, реализованного в виде известного метода группового учета аргументов (МГУА). Для этих исследований разработан оригинальный метод критических дисперсий, позволивший подробно изучить закономерности изменения сложности выбираемой структуры в зависимости от уровня шума. Доказан методологически важный результат о необходимости согласовывать сложность модели с уровнем неопределенности в данных.

Эта часть имеет не только теоретическую, но и прикладную направленность. В ней разрабатываются алгоритмы и программные средства моделирования экономических процессов на основе МГУА. Приведены результаты решения задач моделирования и прогнозирования реальных экономических процессов. Показано, что применение моделей,

построенных по МГУА, для целей прогнозирования дает существенно лучшие результаты по сравнению с моделями, полученными на основе регрессионного анализа.

Здесь разрабатываются также принципы формирования алгоритмического, функционального и диалогового обеспечения программных систем моделирования и прогнозирования сложных процессов на основе МГУА. Эти принципы частично реализованы в разработанном действующем макете проблемно-ориентированной программной системы автоматизированной **структурной** идентификации АСТРИД.

Раздел 1

ЗАДАЧА МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ ПО СТАТИСТИЧЕСКИМ ДАННЫМ

1.1. Моделирование экономических процессов по данным наблюдений в условиях неопределенности как задача структурной идентификации

Для переходной экономики характерна нестабильность экономических отношений, неустойчивость денежно-кредитной системы и недостаточность законодательного регулирования. При этом известные балансовые модели, учитывающие внутри- и межотраслевые соотношения, на практике часто оказываются непригодными для планирования и управления экономическими процессами. В такой ситуации для анализа, моделирования и прогнозирования этих процессов целесообразно применять методы прямого построения моделей по данным наблюдений (государственной статистики). Цель таких методов – выявление неявных причинно-следственных связей и закономерностей, скрытых в ретроспективных данных, и представление их в явной форме математических моделей. При этом необходимо искать как структуру, так и параметры моделей, т.е. решать задачу структурно-параметрической, или просто структурной, идентификации.

Для большинства сложных объектов и процессов управления задача структурной идентификации решается в условиях неопределенности, при которых, как правило, нельзя ставить целью построение модели, имеющей однозначное соответствие с объектом, а можно говорить лишь о поиске *правдоподобной* модели, не противоречащей данным наблюдений и имеющейся априорной информации [1]. При этом вместо термина "идентификация объекта" целесообразно использовать более соответствующий термин "идентификация *модели* объекта". Объем вычислений, необходимых для поиска модели, удовлетворяющей той или иной цели (например, с наилучшими прогнозирующими свойствами), требует применения специализированных программных средств.

Отметим, что с точки зрения задач управления эколого-экономическими процессами целесообразно стремиться получить в результате идентификации модель с наилучшими прогнозирующими свойствами. В связи с этим из всех подходов, существующих в структурной идентификации, наибольший интерес представляет подход "перекрестного подтверждения" (cross-validation). По своей сути он направлен на выбор моделей, уменьшающих ошибку прогнозирования: в простейшем случае выборка делится на две части, на одной из которых генерируется множество моделей, а на второй вычисляется ошибка каждой модели, и выбирается та, которая лучше работает "в режиме прогнозирования". Среди методов этого направления выделяется МГУА [2] как метод автоматического поиска лучшей модели, имеющий большее разнообразие возможностей по сравнению с другими методами. Поэтому в отчете алгоритмы МГУА и опыт их применения приняты в качестве методической основы для сравнительного анализа различных методов и проектирования интерактивных программных средств построения моделей. Это индуктивный метод построения линейных, нелинейных, разностных и других

математических моделей сложных процессов по коротким выборкам данных. В его основу положены принципы внешнего дополнения, автоматической генерации и последовательной селекции усложняющихся структур моделей.

Примеры применения алгоритмов МГУА для моделирования экономик США, Болгарии, ГДР приведены в [3]. Однако для моделирования процессов в переходной экономике Украины эти алгоритмы еще не применялись.

Достаточно общая постановка задачи структурно-параметрической идентификации, или построения моделей по экспериментальным данным, может быть сведена к поиску экстремума некоторого критерия CR на множестве различных моделей \mathfrak{F} :

$$f^* = \arg \min_{f \in \mathfrak{F}} CR(f). \quad (1.1)$$

Очевидно, что (1.1) не содержит исчерпывающей формулировки задачи, поэтому дополнительно необходимо: задать вид и объем исходной информации; указать класс базисных функций (операторов), из которых формируется множество \mathfrak{F} ; определить способ генерации моделей f вместе с методом оценивания параметров; выбрать критерий сравнения моделей; указать метод минимизации CR . Уточним эту постановку, предположив, что задана выборка $W=[Xy]$, содержащая n точек наблюдений, образующих матрицу $X = \{x_{ij}, i=1, \dots, n; j=1, \dots, m\}$ и вектор $y = (y_1 \dots y_m)^T$, причем $n \geq m$.

В общем случае процесс решения задачи структурно-параметрической идентификации (1.1) включает в себя прежде всего следующие основные этапы:

- 1) Задание выборки данных (полученной в результате пассивного или активного эксперимента), а также априорной информации;
- 2) Выбор или задание класса базисных функций и преобразование данных;
- 3) Генерация различных структур моделей в выбранном классе;
- 4) Оценивание параметров генерируемых структур и формирование множества F ;
- 5) Минимизация заданного критерия $CR(f)$ и выбор оптимальной модели f .
- 6) Проверка адекватности полученной оптимальной модели.
- 7) Принятие решения о завершении процесса.

Перечисленные этапы описывают произвольный процесс построения моделей, причем в зависимости от априорной информации и цели моделирования те или иные этапы могут отсутствовать. Например, в случае задачи параметрической идентификации множество F состоит из одной модели (задается одна структура), т.е. исключается этап 5 процесса.

В целом задача идентификации состоит в формировании по данным выборки некоторого множества \mathfrak{F} моделей различной структуры вида

$$\hat{y}_f = f(X, \theta_f) \quad (1.2)$$

и отыскании оптимальной модели по условию

$$f^* = \arg \min_{f \in F} CR(y, f(X, \theta_f)), \quad (1.3)$$

причем оценки параметров в (1.2) для каждой модели $f \in F$ являются решением еще одной экстремальной задачи вида

$$\hat{\theta}_f = \arg \min_{\theta_f \in R^m} QR(y, X, \theta_f), \quad (1.4)$$

где S_f называется сложностью модели f и равно числу ненулевых компонент в модели вида (1.3), а QR - критерий качества решения задачи параметрической идентификации каждой частной модели, генерируемой в задаче структурной идентификации.

Экстремальные задачи (1.3), (1.4) содержат в себе ряд неопределенностей, в принципе свойственных в целом задаче моделирования по данным наблюдений и в той или иной степени влияющих на качество ее решения. Совокупность неопределенностей можно разделить на две основные группы: относящиеся к данным, т.е. к физической информации о моделируемом объекте, и относящиеся к средствам обработки данных, т.е. к применяемой технологии моделирования.

Исходные данные содержат следующие *виды неопределенности*:

- 1) *структурная* неопределенность, означающая неполное знание связей вход-выход и не позволяющая однозначно задать структуру модели (1.2).
- 2) *стохастическая* неопределенность в виде неизвестного характера и уровня шума в выборке данных;
- 3) *информационная* неопределенность или качество выборки; в частности, неопределенным в смысле достаточности является само число наблюдений.

Виды неопределенности, относящиеся к *процедуре моделирования*:

- *функциональная* неопределенность: под независимыми переменными x^j , $j = 1, \dots, m$ в (1.2) могут пониматься некоторые функции от реально измеряемых входных переменных $z = (z_1 \dots z_l)$: $x_j = \varphi_j(z)$; тогда при моделировании появляется неопределенность в выборе того или иного базисного набора функций φ_j , $j = 1, \dots, m$;
- *параметрическая* неопределенность, относящаяся к выбору критерия QR и метода (алгоритма) решения задачи параметрической идентификации (1.4);
- *критериальная* неопределенность, относящаяся к выбору критерия CR и метода (алгоритма) решения основной задачи структурной идентификации (1.3).

Каждый из методов решения задач моделирования следует рассматривать с точки зрения степени учета указанных видов неопределенности. В этом смысле крайние позиции занимают имитационное моделирование и МГУА: первый требует от исследователя снятия всех шести неопределенностей, а второй ориентирован на их автоматический учет. Промежуточное положение между этими двумя методами занимает регрессионный анализ, наиболее популярный у пользователей. Поэтому ниже приводится краткое качественное сравнение методологий регрессионного анализа и МГУА.

Отметим, что МГУА разработан чл.-корр. НАНУ Ивахненко А.Г. и является методом структурно-параметрической идентификации сложных процессов и объектов по данным наблюдений в условиях помех.

1.2 Сравнительная характеристика существующих подходов к построению моделей по данным наблюдений

Рассмотрим коротко наиболее популярные методы (за исключением МГУА), применяемые для решения задачи поиска лучшей регрессии вида (1.3). При этом способы формирования базисного набора функций не затрагиваются.

В рассматриваемых методах множество моделей F формируется с помощью генераторов полного или направленного перебора моделей различной сложности [4]. Полный перебор называется “методом всех регрессий” для сравнения моделей вида (1.2), содержащих от одного до m регрессоров (независимых переменных). При направленном переборе формируется сокращенное число моделей с целью получить результат полного перебора при существенно меньших вычислительных затратах. К ним относятся методы включения (добавление регрессоров по одному, максимально уменьшающему критерий), исключения (начинается процедура с полной модели, затем регрессоры убираются по одному, пока критерий убывает) и шаговой регрессии (комбинация двух предыдущих). Известен также метод ветвей и границ, имеющий наибольшую вероятность получения результата полного перебора [4].

Для оценивания параметров в алгоритмах прикладного регрессионного анализа применяется прежде всего МНК, когда в качестве QR в (1.4) выступает функция потерь (сумма квадратов ошибок):

$$QR(\theta_f) = \|y - \hat{y}_f\|^2 = \|y - X\hat{\theta}_f\|^2. \quad (1.5)$$

МНК является наилучшим методом при выполнении предположения о том, что в измерениях y присутствует независимый шум с нулевым средним и конечной дисперсией, применение которого дает решение задачи оценивания (1.5) в виде

$$\hat{\theta}_f = (X_f^T X_f)^{-1} X_f^T y, \quad (1.6)$$

где X_f - подматрица матрицы X , содержащая столбцы, соответствующие рассматриваемой частной модели $f \in \mathfrak{F}$.

Наиболее популярными критериями сравнения моделей, генерируемых в процессе перебора, являются следующие:

- критерий Фишера (дисперсионное отношение) [2]:

$$\Phi(s_f) = \frac{RSS(s_f) / (n - s_f)}{\|\bar{y} - y\|^2 / (n - 1)}, \quad (1.7)$$

где $RSS(\cdot)$ - остаточная сумма квадратов, \bar{y} - среднее значение выхода;

- финальная ошибка прогнозирования [5]:

$$FPE = \frac{n + s_f}{n - s_f} RSS(s); \quad (1.8)$$

- критерий Маллоуза [6] :

$$C_p = \frac{1}{\sigma^2} RSS(s_f) + 2s_f - n, \quad (1.9)$$

где σ^2 - оценка неизвестной дисперсии;

- критерий Акаике [7]:

$$AIC(s_f) = -2 \ln \varphi(X, \theta_f) + 2s_f, \quad (1.10)$$

где φ - значение функции правдоподобия.

Наиболее обоснованными критериями выбора лучшей модели считаются C_p и AIC , однако для первого необходима достоверная оценка дисперсии шума, а для второго - знание функции распределения, что затрудняет их практическое применение.

Все указанные критерии используют ту или иную корректировку RSS , целью которой является учет эмпирически установленного правила: между сложностью модели и ее точностью должен достигаться компромисс. Более подробный обзор методов и критериев решения задач моделирования по экспериментальным данным имеется в [8].

1.3 МГУА как оригинальный метод автоматизированного построения моделей

МГУА является оригинальным методом решения задач структурно-параметрической идентификации моделей сложных систем, обладающим определенным разнообразием возможностей на всех этапах процесса моделирования по сравнению с методами прикладного регрессионного анализа. Это касается прежде всего генераторов моделей и применяемых критериев качества структур, а также классов моделей (базисных функций).

Рассмотрим структурные особенности алгоритмов МГУА, для чего приведем, с учетом сформулированных в подразделе (1.1) основных этапов моделирования, характеристику различных типов классов моделей, генераторов структур и внешних критериев, применяемых в МГУА.

1.3.1 Классы моделей, применяемые в МГУА

Можно указать следующие основные типы задач, встречающиеся в практике моделирования:

1. построение регрессионных моделей статических объектов;
2. моделирование временных рядов или процессов;
3. моделирование динамических объектов, процессов и систем.

Моделирование статических объектов. Отметим прежде всего, что задача может рассматриваться как статическая и в случае процессов или систем, развивающихся во времени, если динамические эффекты (переходные процессы) пренебрежительно малы. В этом смысле термин “статический объект” объединяет все случаи, когда необходимо построить многомерную регрессионную модель вида

$$y_i = f(y_i) + \xi = Pg(x_i) + \xi_i, \quad (1.11)$$

где $i=1, \dots, n$ - номер точки наблюдения, $g(x)$ - известная m -мерная вектор-функция от MX входных переменных, P - неизвестная $m \times MY$ матрица параметров, ξ_i - независимый

некоррелированный случайный вектор с нулевыми средним и конечной неизвестной дисперсией. При этом, с учетом применения линейной оценки параметров по МНК, задача идентификации P решается независимо для каждой l -й компоненты вектора y , $l=1, \dots, MY$, и можно рассматривать каждое уравнение системы (1.11) в отдельности:

$$y_i^l = p_l^T g(x_i) + \xi_i = \sum_{j=1}^m p_{lj} g_j(x_i) + \xi_i^l. \quad (1.12)$$

В качестве базисных функций $g(x)$ в алгоритмах МГУА применяются прежде всего полиномы:

$$p^T g(x) = \sum_j^m p_j \prod_v^{MX} x_v^{c_{jv}}, \quad (1.13)$$

тогда каждая компонента $g(x)$ является одночленом этого полинома. Если при этом степени c_{jv} могут принимать целые значения $0, 1, \dots, ST$, и учитывается ограничение

$$\sum_v^{MX} c_{jv} = C, \quad C = 0, 1, \dots, ST, \quad (1.14)$$

то полином (1.13) (порядок следования и структура одночленов) однозначно задается числами MX и ST , что удобно для программирования, а число его членов равно

$$m = \prod_{j=1}^{MX} \frac{ST + j}{j}, \quad (1.15)$$

Модели временных рядов. Для моделирования процессов как временных рядов в алгоритмах МГУА применяются полиномиальные, тригонометрические, экспоненциальные, логарифмические и логистические базисные функции. Очевидно, что можно моделировать тренды процессов с учетом наличия только одной “входной” переменной - времени, при этом сохраняется задача вида (1.2), (1.3), где $MX = 1$.

В случае тригонометрических (гармонических) функций модели тренда процесса имеют вид (для отдельного временного ряда):

$$b^T g(t) = a_o + \sum_{j=1}^m (\alpha_j \sin \omega_j t + \beta_j \cos \omega_j t) = a_o + \sum_{j=1}^m b_j \sin(\omega_j t + \psi_j), \quad (1.16)$$

т.е. $g_j(t) = \sin(\omega_j t + \psi_j)$ при заданных ω_j, ψ_j . В [9] описан метод одновременного оценивания частот ω_j и амплитуд $\alpha_j \beta_j$ (или ω_j, ψ_j, b_j) при заданном m .

Кроме указанных классов моделей, для моделирования временных рядов применяется также класс динамических прогнозирующих моделей в виде авторегрессионных зависимостей, связывающих отсчеты процесса в различные моменты времени:

$$y_k = \sum_{i=1}^{LY} a_i y_{k-i} + \xi_k, \quad (1.17)$$

где LY - порядок авторегрессии (число учитываемых запаздываний).

Модели динамических объектов. Третий - наиболее общий - тип задач моделирования динамики решается прежде всего в классе линейных динамических моделей. При этом, кроме автономной части модели вида (1.17), имеется также часть, отражающая влияние вектора внешних независимых воздействий x . Тогда общий вид модели динамики многомерной системы представляется известным векторным уравнением:

$$y_k = \sum_{\alpha=1}^{LY} A_{\alpha} y_{k-\alpha} + \sum_{v=1}^{LX} B_v x_{k-v+1} + \xi_k, \quad (1.18)$$

где LY, LX - число учитываемых прошлых значений (запаздываний) для выходных и входных переменных соответственно, $A_1, \dots, A_{LY}, B_1, \dots, B_{LX}$ - матрицы неизвестных параметров, ξ_k имеет смысл входного процесса типа белого шума. В этих условиях и в предположении использования модели (1.18) для прогнозирования на шаг вперед оценки параметров каждого уравнения оказываются независимыми, так что можно рассматривать модель каждого l -го выхода отдельно, $l = \overline{1, MY}$:

$$\begin{aligned} y_k^l &= \sum_{\alpha=1}^{LY} a_{l\alpha}^T y_{k-\alpha} + \sum_{v=1}^{LX} b_{lv}^T x_{k-v+1} + \xi_k^l = \\ &= \sum_{r=1}^{MY} \sum_{\alpha=1}^{LY} a_{l\alpha}^r y_{k-\alpha}^r + \sum_{j=1}^{MX} \sum_{v=1}^{LX} b_{lv}^j x_{k-v+1}^j + \xi_k^l \end{aligned} \quad (1.19)$$

Очевидно, что из (1.19) можно получить ряд предыдущих классов моделей на основе вектора классифицирующих (ключевых) параметров $\{MY, LY, MX, LX\}$ (см. 4.1).

В каждом алгоритме МГУА имеется блок преобразования данных в соответствии с заданным классом моделей, в результате чего формируется и решается стандартная задача структурно-параметрической идентификации (1.3), где специфика типа исходной задачи уже в явном виде не учитывается. Например, в одномерном случае ($MY=1$, индекс l опускаем) выражения (1.12) и (1.19) приводятся к виду обычной регрессионной модели

$$y_i = (f x_i, \theta) + \xi_i = x_i^T \theta + \xi_i, \quad (1.20)$$

если в (1.12) обозначить $\overset{\Delta}{\theta} = \overset{\Delta}{b}, x = \overset{\Delta}{g}(x)$, а в (1.19) - соответственно $\overset{\Delta}{\theta} = (a_1, \dots, a_{LY}, b_1, \dots, b_{LX})^T$, $x = (y_{k-1}, \dots, y_{l-LX}, x_k, \dots, x_{k-LX+1})^T$ ("номер наблюдения" i и "индекс времени" k эквивалентны). Тогда задача (1.1) состоит в определении оптимальной структуры и значений параметра θ , или - в терминах регрессионного анализа - построении модели с оптимальным подмножеством регрессоров (факторов).

В теоретических исследованиях данной работы (раздел 2) рассматривается обобщенное представление моделей вида (1.20), линейных по параметрам, а в разработанных алгоритмах имеется большинство упомянутых выше классов моделей.

1.3.2 Сравнительный анализ генераторов структур моделей

Большинство известных генераторов структур естественным образом разделяются на две группы - переборные и итерационные, по аналогии с методами оптимизации. Рассмотрим характерные особенности этих методов, отличающихся способами генерации структур моделей и организацией поиска минимума заданного критерия.

Переборные методы. Они известны как в методах регрессионного типа, так и в методах самоорганизации, и предназначены для решения задачи перебором моделей из конечного множества \mathfrak{S} . Все элементы этого множества могут быть вычислены независимо с помощью оценки параметров всех различных моделей. При этом задача (1.3) оказывается аналогичной задаче целочисленного (дискретного) программирования и может быть решена полным или направленным перебором.

Известны различные реализации *полного перебора*, однако они эффективны при относительно небольшом числе аргументов (регрессоров). Так, в случае линейного объекта с m входами число сравниваемых моделей вида

$$y_v = X_v \theta_v, v = 1, \dots, 2^m, \quad (1.21)$$

где десятичному числу v ставится в соответствие двоичное структурное число d_v , равно 2^m , и полный перебор реален примерно до $m = 20$.

Цель методов *направленного перебора* - отыскать результат полного перебора, т.е. глобальный минимум $CR(s)$, при существенно меньших вычислительных затратах. Среди них известны методы включения регрессоров по одному (максимально уменьшающему критерий), исключения их по одному из полной модели, и шаговые (включения - исключения) [4]. Эффективнее в смысле отыскания глобального минимума CR методы ветвей и границ, направленного усложнения на основе принципа неокончательных решений и другие. Они работоспособны при достаточно больших m - около 50.

Примером направленного перебора может быть процедура вида

$$\hat{y}_s^l = (X_{s-1}^i | x_s^j) \theta_s, s = \overline{1, m}, i, l = \overline{1, F_{s-1}}, \quad (1.22)$$

где s - номер этапа (и сложность структуры); F_s - число лучших структур (свобода выбора); j - индекс регрессора, отсутствующего в матрице X_{s-1}^i . При этом просматривается некоторое подмножество наиболее перспективных структур из полного множества всех возможных структур, которое с большой вероятностью содержит и решение задачи полного перебора. Если $F_s = C_m^s$, то перебор становится полным.

Итерационные методы. Они известны только в теории самоорганизации моделей, и все их варианты (как и переборные алгоритмы) носят общее название "метод группового учета аргументов" (МГУА). По принципу действия итерационные методы подобны классу методов оптимизации с помощью последовательных приближений, однако используют принцип неокончательных решений (свободу выбора).

В зависимости от способа организации последовательных приближений итерационные алгоритмы МГУА можно разделить на две основные группы: селекционные (многорядные) и релаксационные. *Многорядные* алгоритмы [2] построены на аналогии с биологической селекцией живых организмов. При этом усложнение моделей от r -го ряда селекции к $(r+1)$ -му происходит за счет попарного "скрещивания" лучших моделей предыдущего ряда, согласно следующему частному описанию:

$$y_l^{r+1} = \alpha_l y_i^r + \beta_l y_j^r, \quad r = 0, 1, \dots; \quad i, j = \overline{1, F}; \quad l = \overline{1, C_F^2}, \quad (1.23)$$

где F - заданная свобода выбора, а при $r=0$ справа учитываются исходные регрессоры. В настоящее время (1.23) для исключения потери исходного базиса дополняется параллельной процедурой вида

$$y_{l+k}^{r+1} = \alpha_{l+k} y_i^r + \gamma_k x_k, \quad i = \overline{1, F}; \quad k = \overline{1, m}. \quad (1.24)$$

В таких алгоритмах процедура усложнения прекращается после начала возрастания значений критерия CR .

Алгоритмы МГУА *релаксационного* типа (называемые также многорядными с вычислением остатков [2]) соответствуют методу последовательных приближений с помощью групповых релаксаций [10]: на каждом ряде (шаге приближения) лучшие модели усложняются только за счет добавления исходных аргументов, а приближаемой величиной служат остатки от предыдущих приближений:

$$y_l^{r+1} = y_i^r + \gamma_l x_k, \quad r = 0, 1, \dots; \quad i = \overline{1, F}; \quad k = \overline{1, m}. \quad (1.25)$$

В этих алгоритмах момент останова решения задачи определяется по факту стабилизации критерия, причем здесь, в отличие от переборных, множество моделей можно описать, но нельзя просмотреть ввиду его бесконечности, а в процессе решения формируется некоторое его подмножество (также бесконечное) [9], зависящее от параметров алгоритма. При этом увеличивается разнообразие генерируемых структур. Их эффективность доказывают многочисленные примеры решения практических задач моделирования [2,3]. В целом итерационные методы работоспособны при больших m - порядка 100. Они позволяют получать как линейные, так и нелинейные модели, причем даже в вырожденных задачах, когда длина выборки $n < m$.

Выражения (1.21) - (1.25) наглядно (хотя и упрощено) представляют четыре основных типа генераторов структур в алгоритмах моделирования, включая МГУА. Все они линейные, однако с учетом преобразования, приведшего к эквивалентной матрице X , могут давать и нелинейные по входам модели.

Отметим, что в прикладном регрессионном анализе имеются только переборные генераторы (методы всех регрессий, шаговой регрессии, ветвей и границ), в которых не используется принцип неокончательных решений (свобода выбора), т.е. алгоритмы МГУА предоставляют более широкие возможности генерации моделей. Итерационные генераторы МГУА являются оригинальными, и само название МГУА ассоциируется прежде всего с ними. В данной работе алгоритмы итерационного типа не исследуются и не применяются, а

приведенный классифицирующий обзор представлен для полноты структуризации соответствующих знаний, которая способствует разработке и реализации обобщенных многофункциональных алгоритмических модулей (см. раздел 4).

1.3.3 Внешние критерии качества моделей

Внешние критерии предназначены для применения в задаче (1.3) и основаны на разделении выборки на две или более частей. Здесь рассмотрим разбиение на три непересекающихся подвыборки (подмножества точек) A, B, C , причем обозначим также $A \cup B = W$. Отметим, что в алгоритмах МГУА применяются четыре группы критериев: точности, согласованности, баланса и динамические.

Приведем вычислительные формулы для основных внешних критериев групп точности и согласованности, предполагая разбиение заданным и опуская индекс частной модели. Для стандартизации записи формул примем следующие обозначения: оценка параметров по МНК на некоторой подвыборке G равна

$$\theta_G = (X_G^T X_G)^{-1} X_G^T y_G, \quad G = A, B, W, \quad (1.26)$$

а значение ошибки на некоторой подвыборке Q по модели, оценки параметров которой вычислены на G , равно

$$\Delta(Q|G) = \|y_Q - X_Q \theta_G\|^2, \quad (1.27)$$

где $Q = A, B, W, C$. Разбиение на три подвыборки соответствует следующему:

$$X = \begin{bmatrix} X_A \\ X_B \\ X_C \end{bmatrix} = \begin{bmatrix} X_W \\ X_C \end{bmatrix}, \quad Y = \begin{bmatrix} Y_A \\ Y_B \\ Y_C \end{bmatrix} = \begin{bmatrix} Y_W \\ Y_C \end{bmatrix}. \quad (1.28)$$

Критерии точности. Очевидно, что при $Q = G$ величина $\varepsilon = \Delta(G/G)$ является остаточной суммой квадратов RSS , а при $G \neq Q$ она является критерием *регулярности*:

$$AR_B = \Delta(B|A), \quad AR_A = \Delta(A|B). \quad (1.29)$$

Тогда критерий *двойной (симметричной) регулярности*:

$$AD = \Delta(B|A) + \Delta(A|B). \quad (1.30)$$

Критерий *стабильности* (на W) определяется выражением:

$$AS = \Delta(W|A) + \Delta(W|B) = AD + \varepsilon_A + \varepsilon_B. \quad (1.31)$$

Критерии согласованности. Основным критерием этой группы является критерий *непротиворечивости* (несмещенности), определяемый выражениями:

$$\begin{aligned} CB &= \|X_W \theta_A - X_W \theta_B\|^2 = (\theta_A - \theta_B)^T X_W^T X_W (\theta_A - \theta_B) = \\ &= AS - 2(\varepsilon_A + \varepsilon_B). \end{aligned} \quad (1.32)$$

К данной группе относится также критерий *вариативности*

$$CV = (\theta_A - \theta_W)^T X_W^T X_W (\theta_W - \theta_B) = (\varepsilon_W - (\varepsilon_A + \varepsilon_B)). \quad (1.33)$$

Описанные критерии предполагают один вариант разбиения выборки. Имеются также критерии с многократным разбиением выборки - например, JN (джек-наиф или “усредненный критерий регулярности” (УКР)). Подробнее о критериях см. в [11].

1.4 Качественное сравнение МГУА и других статистических методов

МГУА и методы прикладного регрессионного анализа можно сравнивать по следующим аспектам, существенным для успешного решения задач вида (1.1) или (1.3):

- возможности автоматической генерации различных структур моделей в заданном классе структур;
- методы оценивания параметров генерируемых структур;
- применяемые критерии для сравнения и выбора моделей оптимальной сложности;
- способы оценки адекватности получаемых моделей;
- работоспособность при различных нарушениях основных предположений;
- разнообразие применяемых классов базисных функций;
- степень автоматизации процессов моделирования (“комфортность” пользователя);
- работоспособность при малых и больших выборках данных.

Сравнительные особенности алгоритмов МГУА по указанным аспектам в нестрогом изложении сводятся к следующему:

1. МГУА представляет значительно более широкие возможности автоматической генерации структур моделей. Генераторы МГУА есть двух типов: переборные и итерационные. Переборные делятся на комбинаторные (аналоги метода всех возможных регрессий) и многоэтапные (аналоги методов шаговой регрессии и ветвей и границ, однако с существенным дополнением в виде применения принципа неокончательных решений). Итерационные генераторы (аналогов в регрессионном анализе не имеют) делятся на селекционные (имитируют процедуру биологической селекции с попарным учетом последовательных признаков) и релаксационные (с последовательным моделированием остатков от предыдущих приближений). Алгоритмы с итерационным усложнением моделей предоставляют принципиально новые возможности в задаче приближения функций и неизвестных зависимостей (разложение в “ряд МГУА”).

2. Для оценивания параметров отдельных моделей в МГУА применяется МНК, и информация (если она имеется) о характеристиках помех в настоящее время прямо не используется. Использование такой информации (на практике она редко имеется в распоряжении) возможно - для этого следует изменить соответствующий модуль алгоритма. Это, по существу, имеет значение только для переборных алгоритмов, поскольку в итерационных в процессе решения строятся специфические “оценки МГУА”, которые не совпадают с МНК-оценками даже при структурном совпадении моделей.

3. Для сравнения и выбора моделей в МГУА применяются внешние критерии, основанные на разбиении выборки на две и более частей, причем проверка качества модели и оценивание параметров по этим критериям выполняется на различных подвыборках. Это позволяет обойтись без предположений о свойствах шумов и уровнях доверия (критерий

Фишера), а также без оценивания дисперсии шумов или значений функции распределения (критерии Маллоуза, Акаике). При этом в условиях нарушений основных предпосылок достигается автоматический учет этих нарушений (которые отражены в данных), за счет чего получается эффект помехоустойчивого моделирования ввиду выбора модели оптимальной сложности (в смысле учета неполноты данных и зашумленности выборки).

4. Оценка адекватности в алгоритмах МГУА выполняется по значениям внешних критериев, а также по величине точности экстраполяции (прогнозирования) на независимой выборке данных. При выполнении основных предположений можно вполне применять традиционные методы оценки адекватности (Фишера, Колмогорова и др.).

5. Алгоритмы МГУА можно отнести к робастным методам моделирования в смысле выбора модели с теоретически минимальной дисперсией прогноза (за счет сокращения сложности модели ввиду неявного учета уровня шум/сигнал и различных нарушений - эффект помехоустойчивости). При этом структура модели адаптируется к объему информации, заложенной в выборке (чем меньше информации, тем проще модель).

6. Разнообразие применяемых классов функций в подходах в принципе одинаково; при этом в алгоритмах МГУА имеется много различных автоматически генерируемых вариантов классов моделей.

7. Алгоритмы МГУА обладают высокой степенью автоматизации процесса моделирования - достаточно ввести данные и указать внешний критерий; однако качество моделей при этом связано с опытом выбора управляющих параметров алгоритмов.

8. При малых выборках данных алгоритмы МГУА обладают преимуществом за счет выбора сложности модели, оптимально учитывающей информативность данных. При возрастании длины выборки два подхода в конце концов оказываются теоретически эквивалентными (эффект состоятельности критериев МГУА [12]).

1.5. Многоэтапный (комбинаторно-селекционный) алгоритм с полиномиальной сложностью перебора моделей

Так как объем полного перебора является *показательной* (экспоненциальной) функцией числа регрессоров, то очевидна ограниченность возможностей комбинаторных алгоритмов. Поэтому для решения более сложных задач необходимо разрабатывать способы сокращенного перебора.

В данном параграфе описан новый алгоритм сокращения полного перебора, являющийся, как и многорядные алгоритмы МГУА, селекционным, однако имеющий конечное число этапов (не более m) и позволяющий искать модель в том же заданном функциональном базисе, который используется в переборных алгоритмах. Предлагаемая схема сокращенного перебора отличается от известных шаговых схем прежде всего применением принципа неокончательных решений, существенно повышающего вероятность получения результата полного перебора.

Общая *схема алгоритма*: сначала оцениваются все модели с одним аргументом, и выбирается несколько лучших из них; на втором этапе к этим моделям добавляются

различные аргументы по одному, и отбираются модели с двумя аргументами, улучшающие величину критерия, и т.д. до его ухудшения. Такой алгоритм можно назвать *многоэтапным* в отличие от многорядных, где число рядов может быть бесконечным.

Последовательность операций, выполняемых на каждом s -м этапе алгоритма:

1. Генерируются частные модели s -го этапа упорядоченным добавлением к каждой i -й модели предыдущего этапа, $i = \overline{1, F_{s-1}}$, одного из отсутствующих в ней аргументов, соответствующих нулевым элементам структурного вектора d_i i -й модели. То есть на этапе генерируется максимум $q_s = (m-s)F_{s-1}$ частных моделей со сложностью (числом аргументов или оцениваемых параметров) s , причем q_s является завышенной оценкой, включающей возможные повторения структур, которые несложно алгоритмически избежать. Каждая l -я новая модель, $l = \overline{1, q_s}$, однозначно определяется номером i модели $(s-1)$ -го этапа и номером j добавленного аргумента, $j = \overline{1, m-s}$ (см. формулу (1.22)).

3. Вычисляются оценки коэффициентов $\hat{\theta}_s(i, j)$ при каждом i, j по МНК.

4. Вычисляются значения критерия селекции для каждой из q_s генерируемых моделей.

5. Определяются FS_s лучших моделей, передаваемых на следующий этап селекции, по условию улучшения значений критериев CR_{s-1}^i , полученных на предыдущем этапе для каждой i -й модели:

$$FS_s = \sum_{i=1}^{FS_{s-1}} \sum_{j=1}^{m-s} \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & CR_s^i(j) < CR_{s-1}^i, \\ 0, & CR_s^i(j) \geq CR_{s-1}^i. \end{cases} \quad (1.34)$$

Число FS_s при больших m может оказаться слишком большим для конкретной ЭВМ, поэтому необходимо ввести некоторое целесообразное ограничение сверху FM . Тогда свобода выбора на s -м этапе определяется по (1.34) с учетом ограничения:

$$FS_s \leq FM.$$

6. Принимается решение об останове процедуры или переходе на этап 1. Очевидно, что условием останова является $FS_s = 0$, и модель оптимальной сложности определяется минимальным значением критерия на одном из предыдущих этапов.

Описанный алгоритм назван MULTI. Его *основные свойства*:

1. Алгоритм является селекционным, с конечным числом этапов перебора, причем свобода выбора на каждом этапе и момент останова определяются автоматически.

2. Все анализируемые модели принадлежат тому же функциональному базису, что и модели обычного комбинаторного алгоритма.

3. На каждом этапе сложность частных моделей увеличивается только на один аргумент. При этом затраты памяти оказываются теми же, что и в COMBI.

4. Алгоритм позволяет получить некоторые известные схемы перебора: при $FM = C_m^s$ - полный перебор (метод всех регрессий), при $FM = 1$ - последовательный перебор (метод включения).

5. Справедлива следующая оценка сверху общего числа моделей, сравниваемых по этому алгоритму:

$$\bar{q}_m = \sum_{s=1}^m q_{smax} = \sum_{s=1}^m (m-s) FM = \frac{1}{2} m(m-1) FM \quad (1.35)$$

Таким образом, число проверяемых моделей \bar{q}_m является **степенной** функцией числа аргументов m вместо *показательной* зависимости 2^m при полном переборе, т.е. алгоритм работоспособен и при больших m ввиду $\bar{q} \ll P_m$. Например, при $m = 100$ и $FM=25$ оценивается не более $\bar{q}_m = 123750$ моделей, что примерно эквивалентно объему полного перебора при $m=17$ ($\bar{q}_{100} < P_{17}$). То есть алгоритм имеет **полиномиальную** сложность.

6. Если истинные аргументы более коррелированы с выходом y , чем избыточные, алгоритм обеспечивает получение результата полного перебора, поскольку селекционная процедура поиска выбирает наиболее вероятные пути усложнения моделей. В более сложных случаях вероятность достижения результата полного перебора остается достаточно высокой, что объясняется, кроме указанного, также следующими факторами: каждая усложняемая модель проверяется на возможность включения в нее отсутствующих в ней аргументов; для проверки единственности решения процедура останавливается только при $FS=0$; после получения истинной модели (или модели оптимальной сложности при зашумленных данных) алгоритм гарантированно останавливается.

Исследование эффективности комбинаторно-селекционного алгоритма MULTI. Вычислительная эффективность алгоритма исследовалась экспериментально на примере расчета модели экосистемы Каховского водохранилища. Расчеты моделей проводились параллельно по двум алгоритмам – MULTI и COMBI. По алгоритму COMBI, обеспечивающему полный перебор возможных моделей для данного состава аргументов, определялась оптимальная модель в виде линейного полинома со свободным членом. Затем для того же состава аргументов выполнялось моделирование по алгоритму MULTI. Расчеты выполнялись на компьютере РС АТ с процессором 80386.

В качестве исходных данных при моделировании были использованы среднесезонные значения 13 входных переменных за 8 лет, т.е. по 32 значения каждой переменной. Вместе с входными переменными вводились сформированные дополнительные аргументы в виде запаздывающих значений выходной и пяти входных переменных. Количество аргументов формируемых полиномов L изменялось от 7 до 19.

Значения общих для обоих алгоритмов рабочих параметров при моделировании задавались одинаковыми, причем для каждого значения L отбирались 3 лучшие модели.

Исследование эффективности алгоритма MULTI проводилось в два этапа. На первом этапе расчеты по этому алгоритму выполнялись при наибольших возможных значениях параметров SA и FM : $SA=FM=L+1$. На втором этапе были установлены оптимальные значения этих параметров для каждого задаваемого значения L .

Результаты сравнения двух алгоритмов приведены в таблицах 1 и 2. В таблице 1 показано, какое количество моделей для заданного числа аргументов L проверяется при

работе каждого из этих алгоритмов (N_{combi} и N_{multi}) для получения одинакового результата (полного совпадения трех лучших моделей). В таблице 2 приведено время счета по каждому алгоритму для разных значений L (T_{combi} и T_{multi}). Результаты первого этапа исследования, когда $SA=FM=L+1$, приведены в третьих строчках таблиц 1 и 2 (N_{multi1} и T_{multi1}), а результаты второго этапа - в четвертых строчках (N_{multi2} и T_{multi2}).

Оптимальное значение параметра SA (число этапов алгоритма) определялось по результатам работы алгоритма COMBI (наибольшее количество аргументов в каждой из 3 отобранных моделей). Это значение SA может быть определено при работе самого алгоритма MULTI, когда достигаются наилучшие значения критерия отбора для трех моделей и дальнейший перебор моделей не приводит к их улучшению. Оптимальное значение параметра FM определялось экспериментально путем перебора. Вопрос выбора этой величины будет рассмотрен в дальнейших исследованиях.

Как видно из приведенных результатов, при наибольших возможных значениях параметров $SA=FM=L+1$ количество проверяемых моделей по алгоритму MULTI при $L=7$ меньше в 2 раза, а при $L=14$ – в 34 раза. Так же сокращается и время счета.

При оптимальных значениях параметров SA и FM перебор по алгоритму MULTI при $L=14$ сокращается в 125 раз, а время счета – в 175 раз. Соответствующие графики, характеризующие быстроедействие MULTI, показаны на рисунке 1.1.

Таблица 1.1 – Количество сравниваемых моделей

L	7	8	9	10	11	12
N_{combi}	256	512	1024	2048	4096	8192
N_{multi1}	125	184	264	357	474	626
N_{multi2}	53	69	82	95	108	195

Продолжение таблицы 1.1

L	13	14	15	16	17	18	19
N_{combi}	16384	32768	65536	131072	262144	524288	1048576
N_{multi1}	762	964	1212	1483	1790	2198	2711
N_{multi2}	186	261	366	446	498	875	1120

Таблица 1.2 – Время счета

L	7	8	9	10	11	12
T_{combi}	7''	15''	35''	1'20''	2'55''	6'32''
T_{multi1}	5''	7''	10''	16''	23''	32''
T_{multi2}	1''	2''	2''	3''	4''	8''

Продолжение таблицы 1.2

L	13	14	15	16	17	18	19
T_{combi}	13'50''	32'	—	—	—	—	—
T_{multi1}	42''	55''	1'16''	1'44''	2'21''	3'11''	4'25''
T_{multi2}	7''	11''	16''	23''	27''	50''	1'14''

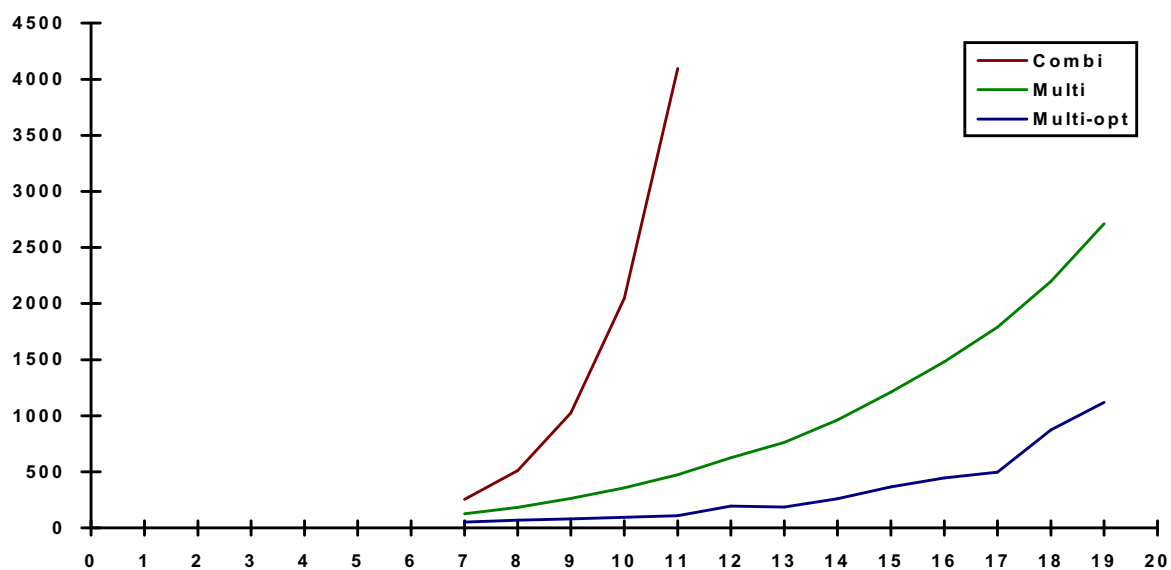


Рисунок 1.1 – Сравнение быстродействия алгоритмов полного перебора моделей и комбинаторно-селекционного (многоэтапного) алгоритма направленного перебора

РАЗДЕЛ 2

ИССЛЕДОВАНИЕ ЗАКОНОМЕРНОСТЕЙ ВЫБОРА СТРУКТУРЫ МОДЕЛИ ПО УСЛОВИЮ МИНИМУМА ДИСПЕРСИИ ОШИБКИ

2.1 Постановка задачи структурной идентификации моделей с минимальной дисперсией ошибки

Рассмотрим постановку задачи, приняв следующие основные предположения:

П1. Объект или процесс имеет m детерминированных входных (независимых) переменных x_1, \dots, x_m и одну выходную (зависимую) переменную y , связанные линейно:

$$y_i = x_i^T \theta + \xi_i = \overset{o}{y} + \xi_i, \quad \overset{o}{y} = x_i^T \theta, \quad (2.1)$$

где θ - вектор m неизвестных параметров, причем из них $s_0 \leq m$ ненулевые, i - номер наблюдения (во времени или пространстве), ξ - случайная величина (шум).

П2. Случайная величина имеет свойства:

$$E \xi_i = 0, \quad E \xi_i^2 = \sigma^2, \quad E \xi_i \xi_j = 0, \quad i \neq j, \quad (2.2)$$

где σ^2 - конечная неизвестная дисперсия, причем функция распределения ξ_i не известна; E - оператор математического ожидания по всем возможным реализациям случайной величины. В соответствии с этим в (2.1) имеем $\overset{o}{y} = Ey$.

П3. Исходная информация задана выборкой данных n наблюдений в виде матрицы $W = [X|y]$, $\dim W = n \times (m + 1)$, где $X = \{x_{kj}, k = \overline{1, n}; j = \overline{1, m}\}$ - матрица регрессоров полного ранга $\text{rank } X = m$, а $y = (y_1 \dots y_n)^T$ - выходной вектор. Кроме того, будем считать заданными также n_F точек, характеризующих область прогнозирования, в виде матрицы X_F , $\dim X_F = n_F \times m$, причем в общем случае $X_F \neq X$.

П4. В асимптотике регрессоры будем считать сильно регулярными [13]:

$$\lim_{n \rightarrow \infty} \frac{1}{n} X_n^T X_n = \overline{H}, \quad (2.3)$$

где \overline{H} - невырожденная конечная $m \times m$ матрица.

Первое допущение означает существование *точной* (или истинной, или физической) модели объекта вида

$$\overset{o}{y}_k = \theta_o^T \overset{o}{x}_k, \quad (2.4)$$

где $\overset{o}{x} \in R^{s_o}$, причем все $\overset{o}{x}^j$, $j = \overline{1, \dots, s_o}$, являются частью компонент измеряемого вектора x , а θ_o - истинный вектор параметров, т.е. компоненты вектора θ в (2.1) равны или компонентам вектора θ_o , или нулю, но при этом ни s_o , ни значения θ_o^j , $j = \overline{1, \dots, s_o}$, ни порядок их размещения в векторе θ не известны. Тогда задача структурной идентификации должна состоять в определении наилучшего в некотором смысле

приближения к неизвестным значениям s_o, θ_o , при одновременном разделении входных переменных x на истинные $\overset{o}{x}$ и избыточные \tilde{x} , $\dim \tilde{x} = (m - s_o) \times 1$.

Для более точной формулировки задачи вида (1.1) рассмотрим, как при принятых предположениях формируется множество \mathfrak{S} моделей-претендентов на лучшую модель.

Определение 1.1. Структурным вектором назовем m -мерный вектор $d = \{d^1, \dots, d^m\}$, состоящий из s_d единиц и $m - s_d$ нулей, $s_d = \overline{0, m}$, указывающих на наличие (при $d^j = 1$) или отсутствие (при $d^j = 0$) в проверяемой модели соответствующей компоненты входного вектора x ; при этом число s_d называется сложностью модели.

Количество неповторяющихся вариантов размещения s единиц в m -мерном двоичном векторе равно числу комбинаций $C_m^s = \frac{m!}{s!(m-s)!}$, поэтому общее число всевозможных структурных векторов равно

$$q_m = \sum_{s=0}^m C_m^s = 2^m, \quad (2.5)$$

причем сюда включен случай $s = 0$, соответствующий независимости выхода от какого-либо из измеряемых входов. Тогда структурный вектор d принадлежит множеству D различных m -мерных бинарных векторов: $D = \{d_l \in \{0, 1\}^m, l = 1, \dots, 2^m\}$.

Определение 1.2. Функция $f(d, \theta, x) = x^T \theta(d) = x^T (\text{diag } d) \theta$ (где $\text{diag } d$ - диагональная матрица с элементами, равными $d^j, j = \overline{1, m}$), заданная с точностью до параметров, называется *структурой модели*, или просто *структурой*.

При заданном входном векторе x структура однозначно задается структурным вектором $d \in D$, т.е. числом и размещением единиц в d , поэтому вместо выражения "структура модели, соответствующая вектору d " будем говорить просто "структура d ".

Допустим теперь, что задан некоторый унимодальный критерий качества оценивания параметров $QR(\theta_d, d, W)$ на W , и определим оценку $m \times 1$ -вектора $\theta(d)$ через оценку $\hat{\theta}_d$ вектора θ_d размерности $d_d \times 1$, составленного из ненулевых компонент $\theta(d)$:

$$\hat{\theta}_d = \arg \min_{\theta_d \in R^{s_d}} QR(\theta_d, d, X, y). \quad (2.6)$$

Тогда множество \mathfrak{S} сравниваемых в задаче (1.1) моделей принимает вид

$$\mathfrak{S} = \{f(d, \hat{\theta}_d, x), d \in D\}, \quad (2.7)$$

где модель $f(d, \hat{\theta}_d, x) = \hat{\theta}_d^T x(d)$, а вектор $x(d)$ размерности $s_d \leq m$ состоит только из компонент вектора x , соответствующих единицам в d .

Критерий $CR(f)$ в (1.1) неотрицательный и отражает некоторую цель моделирования. При этом $CR(f) = CR(d, \hat{\theta}_d, W)$, и в общем случае $CR(d, \hat{\theta}_d, W) \neq QR(d, \hat{\theta}_d, W)$, то есть при

структурно-параметрической идентификации задача дискретной оптимизации (1.1) решается совместно с задачей непрерывной оптимизации (2.6). Результатом решения является оптимальная модель $f(d^*, \theta^*, x) = (\theta^*)^T x(d^*)$.

Здесь остался не определенным способ формирования структурного вектора d , который задает собой одновременно и процедуру поиска экстремума $CR(\cdot)$ на множестве (2.7). Для уточнения рассмотрим следующую задачу поиска лучшего структурного вектора:

$$d^* = \operatorname{argmin}_{d \in D} CR(d, \theta_d, W). \quad (2.8)$$

Всякому $d \in D$ однозначно соответствует θ_d как решение (2.6), и задача (2.8) эквивалентна (1.1), поэтому (2.8) будем рассматривать в качестве задачи структурной идентификации. Она является задачей полного перебора 2^m вариантов структур моделей, т.е. является комбинаторной, с *показательным* относительно m ростом числа вариантов. Эту задачу можно также представить в эквивалентном виде

$$d^* = \operatorname{argmin}_{s=1, m} \min_{d \in D_s} CR(d, \theta_d, W), \quad (2.9)$$

где $D_s \in D$ - подмножество, содержащее C_m^s структурных векторов из s единиц, и $\bigcup_{s=1}^m D_s = D$. Выражения (2.8) и (2.9) – это два варианта полного перебора структур.

Однако (2.9) определяет собой также целый класс алгоритмов направленного поиска с сокращением перебора с помощью тех или иных способов формирования при каждом s некоторых подмножеств $D'_s \subset D$ (см.[4], а также подраздел 1.5). В простейшем случае D'_s содержит один элемент, и (2.9) соответствует последовательному включению переменных в модель, начиная с первой. При этом оптимальная структура модели ищется в так называемом множестве “вложенных” структур [14], тогда вместо (2.9) правомерно рассматривать задачу определения оптимального порядка, или *оптимальной сложности* модели:

$$s^* = \operatorname{argmin}_{s=1, m} CR(s, \theta_s, W). \quad (2.10)$$

Далее в исследованиях будем рассматривать именно этот вариант задачи структурной идентификации, обозначая X_s, θ_s с размерностями $n \times s, s \times 1$ соответственно. Это не ограничивает общность задачи: X_s является произвольным подмножеством $s \leq m$ регрессоров (столбцов матрицы X), а последовательность $s = 1, \dots, m$ их включения в модель - произвольным путем усложнения структуры. При этом каждому s однозначно соответствует оценка $\hat{\theta}_s$, и обозначим $CR(s) \stackrel{\Delta}{=} CR(s, \hat{\theta}_s, W), QR(s) \stackrel{\Delta}{=} QR(s, \theta_s, W)$.

Поскольку ввиду П2 наилучшим оценителем является МНК [13], т.е.

$$QR(s) \stackrel{\Delta}{=} RSS(s) = \|y - X_s \hat{\theta}_s\|^2,$$

где $\|\cdot\|$ - евклидова норма, а $RSS(s)$ - остаточная сумма квадратов, решением задачи (2.7) является оценка МНК:

$$\hat{\theta}_s = (X_s^T X_s)^{-1} X_s^T y.$$

Рассмотрим, наконец, вопрос о критерии $CR(s)$, который следует назначать с учетом основной цели моделирования - а таковой нами принят выбор модели с наилучшими прогнозирующими свойствами. Для формализации этого воспользуемся высказанной в [15] идеей об аналогии задач моделирования и теории связи [16,17]. Зашумленный выход объекта y можно рассматривать как сигнал после канала с шумом, содержащий истинный

сигнал $\overset{o}{y}$, скрытый в шумах. По этой выборке можно вычислить оценки

$\hat{y}_s = X_s \hat{\theta}_s, s = \overline{1, m}$, каждая из которых является оценкой сигнала $\overset{o}{y}$ по одному из алгоритмов обработки зашумленного сигнала y , которыми обладает приемник. Тогда, согласно теории оптимального приемника Котельникова [16], наилучшим алгоритмом является тот, который минимизирует так называемый "идеальный критерий"

$$J(s) \overset{\Delta}{=} J(s, X) = E \left\| \overset{o}{y} - \hat{y}_s \right\|^2 = E \left\| \overset{o}{y} - X_s \hat{\theta}_s \right\|^2, \quad (2.11)$$

Он характеризует качество восстановления истинного сигнала $\overset{o}{y}$ по зашумленному y и равен суммарной дисперсии ошибки восстановления сигнала в точках выборки.

Аналогично (2.11), качество прогнозирования в точках X_F по модели сложности s , коэффициенты которой $\hat{\theta}_s$ получены на обучающей выборке $W=[Xy]$, естественно характеризовать величиной ожидаемых потерь в области прогноза X_F , или величиной суммарной дисперсии ошибки прогнозирования:

$$J_F(s) \overset{\Delta}{=} J_F(s, X, X_F) = E \left\| \overset{o}{y}_F - \hat{y}_{Fs} \right\|^2 = E \left\| \overset{o}{y}_F - X_{Fs} \hat{\theta}_s \right\|^2, \quad (2.12)$$

Очевидно, что в случае $X_F = X$ (2.12) превращается в (2.11).

Определение 1.5. Структуры моделей со сложностями

$$s_F^o = \arg \min_{s=1, m} J_F(s), \quad (2.13)$$

$$s^o = \arg \min_{s=1, m} J(s), \quad (2.14)$$

называются соответственно J_F -оптимальной (оптимальной прогнозирующей) и J -оптимальной (оптимальной восстанавливающей) структурами.

Отметим, что алгоритм, минимизирующий (2.11), в теории идеального приемника Котельникова называется потенциально (предельно) помехоустойчивым, поэтому структуру сложности s^o будем называть также потенциально помехоустойчивой для восстановления, а s_F^o - соответственно для прогнозирования истинного (незашумленного) сигнала. Критерий типа $J(s)$ впервые был введен Маллоузом [6] как величина дисперсии ошибки "предсказания" (prediction) сигнала, поэтому мы также будем различать термины "прогнозирование" и "предсказание" (т.е. восстановление).

Так как функционалы $J_F(s)$, $J(s)$ не могут быть вычислены по данным выборки, возникает проблема их оценивании с помощью каких-либо вычислимых критериев, которые рассмотрены в подразделах 1.2 и 1.3.

2.2. Метод критических дисперсий в задаче структурной идентификации моделей

Вычислим математические ожидания от теоретической ошибки прогнозирования:

$$R_F(s) = \left\| y_F^o - X_{Fs} \hat{\theta}_s \right\|^2 = (y_F^o - X_{Fs} \hat{\theta}_s)^T (y_F^o - X_{Fs} \hat{\theta}_s), \quad (2.15)$$

С учетом принятых выше предположений получаем:

$$E[\hat{\theta}_s^T X_{Fs}^T X_{Fs} \hat{\theta}_s] = \bar{\theta}_s^T X_{Fs}^T X_{Fs} \bar{\theta}_s + \text{tr}(X_{Fs}^T X_{Fs})^{-1} X_{Fs}^T X_{Fs},$$

где $\text{tr } H$ - след матрицы, $\bar{\theta}_s = E[\hat{\theta}_s]$.

Тогда формула (2.12) принимает вид:

$$J_F(s) = J_F^b(s) + J_F^v(s) = \left\| y_F^o - X_{Fs} \bar{\theta}_s \right\|^2 + \sigma^2 \text{tr}(X_{Fs}^T X_{Fs})^{-1} X_{Fs}^T X_{Fs}, \quad (2.16)$$

где индексом "b" (от "bias") отмечена величина средних потерь $J_F^b(s)$ от смещения структуры модели, а индексом "v" (от "variation") - величина вариации функционала, или средние дополнительные потери $J_F^v(s)$ вследствие наличия шума. В дальнейшем $J_F^b(s)$ будем называть "структурная составляющая", а $J_F^v(s)$ - "шумовая составляющая".

Выражение для (2.11) несложно получить из (2.16) при условии $X_F = X$:

$$J(s) = J^b(s) + J^v(s) = \left\| y^o - X_s \bar{\theta}_s \right\|^2 + \sigma^2 s. \quad (2.17)$$

Очевидно, что составляющая $J^b(s)$ является монотонно убывающей функцией s (ввиду уменьшения ошибки МНК при увеличении числа членов модели), а $J^v(s)$ линейно возрастает с ростом s . Следовательно, компоненты $J^b(s)$ и $J^v(s)$ противоречивы, и функция $J(s)$ как сумма этих двух компонент всегда имеет минимум в некоторой точке $s^o \in [1, m]$.

Получим рекуррентные по s выражения для $J^b(s), J^v(s)$, основываясь на известной формуле вычисления оценки параметров регрессионной модели при добавлении одной переменной (регрессора) [4]: если $\bar{\theta}_s$ является МНК-оценкой для модели, содержащей s переменных $X_s = (x_1, \dots, x_s)$, то при добавлении $(s+1)$ -го регрессора параметр $\bar{\theta}_{s+1}$ модели с регрессорами $X_{s+1} = (X_s | x)$ можно вычислять по рекуррентной формуле

$$\bar{\theta}_{s+1} = \begin{bmatrix} \bar{\theta}_s^* \\ \bar{\omega}_{s+1} \end{bmatrix} = \begin{bmatrix} \bar{\theta}_s - \bar{\omega}_{s+1} (X_s^T X_s)^{-1} X_s^T x \\ \bar{\omega}_{s+1} \end{bmatrix}, \quad (2.18)$$

где $(s+1)$ -я компонента вектора $\bar{\theta}_{s+1}$ равна

$$\bar{\omega}_{s+1} = \frac{\alpha_{s+1}}{\beta_{s+1}}, \quad \alpha_{s+1} = x^T D_s^o y, \quad \beta_{s+1} = x^T D_s x, \quad (2.19)$$

причем $D_s = I_n - X_s (X_s^T X_s)^{-1} X_s^T$ - идемпотентная матрица: $D_s^2 = D_s$, и $\beta_{s+1} > 0$ для всех $s = 0, 1, \dots, m-1$ ввиду $\text{rank } X = m$.

С учетом (2.18), (2.19) несложно получить рекуррентные выражения:

$$J^b(s+1) = J^b(s) - \bar{\omega}_{s+1} \beta_{s+1} = J^b(s) - \alpha_{s+1}^2 / \beta_{s+1}, \quad (2.20)$$

$$J^v(s+1) = J^v(s) + \alpha^2, \quad (2.21)$$

Видно, что $J^b(s)$ строго монотонно убывает, а $J^v(s)$ линейно возрастает.

Получим теперь соответствующие формулы для $J_F^b(s)$ и $J_F^v(s)$. Для $J_F^b(s+1)$ имеем (здесь $X_{F,s+1} = (X_{Fs} | x_F)$):

$$\begin{aligned} J_F^b(s+1) &= \left\| \begin{matrix} o \\ y - X_{F,s+1} \bar{\theta}_{s+1} \end{matrix} \right\|^2 = \left\| \begin{matrix} o \\ y_F - X_{Fs} \bar{\theta}_s^* - \bar{\omega}_{s+1} x_F \end{matrix} \right\|^2 = \\ &= \left\| \begin{matrix} o \\ (y_F - X_{Fs} \bar{\theta}_s) - \bar{\omega}_{s+1} (x_F - X_{Fs} (X_s^T X_s)^{-1} X_s^T x) \end{matrix} \right\|^2 \end{aligned}$$

следовательно,

$$J_F^b(s+1) = J_F^b(s) - 2\bar{\omega}_{s+1} a_F^T b_F + \bar{\omega}_{s+1}^2 b_F^T b_F, \quad (2.22)$$

где введены обозначения

$$a_F = a_F(s) = \begin{matrix} \Delta \\ y - X_{Fs} \bar{\theta}_s \end{matrix}, \quad b_F = b_F(s) = \begin{matrix} \Delta \\ x_F - X_{Fs} (X_s^T X_s)^{-1} X_s^T x \end{matrix}. \quad (2.23)$$

Отметим, что при $X_F = X$ (2.23) превращается в (2.20).

Рекуррентное выражение для $J_F^s(s)$ получим с помощью формулы Фробениуса для обращения блочных матриц.

На ее основе имеем:

$$(X_{s+1}^T X_{s+1})^{-1} = \left[\begin{array}{c|c} X_s^T X_s & X_s^T x \\ \hline x^T X & x^T x \end{array} \right]^{-1} = \left[\begin{array}{c|c} H_s^{-1} + \frac{1}{\beta_{s+1}} c_s c_s^T & -\frac{1}{\beta_{s+1}} c_s \\ \hline -\frac{1}{\beta_{s+1} c_s^T} & \frac{1}{\beta_{s+1}} \end{array} \right],$$

где $H = X_s^T X_s$, β_{s+1} определено в (2.10), а $c_s = H_s^{-1} X_s^T x$. Тогда

$$\begin{aligned} \text{tr}(X_{s+1}^T X_{s+1})^{-1} X_{F,s+1}^T X_{F,s+1} &= \text{tr}(X_s^T X_s)^{-1} X_{Fs}^T X_{Fs} + \\ &+ \frac{1}{\beta_{s+1}} (x^T X_s H_s^{-1} X_{Fs}^T X_{Fs} H_s^{-1} X_s^T x - 2x^T X_s H_s^{-1} X_{Fs}^T x_F + x_F^T x_F) \end{aligned}$$

из чего следует искомое соотношение:

$$J_F^v(s+1) = J_F^v(s) + \frac{\sigma^2}{\beta_{s+1}} b_F^T b_F, \quad (2.24)$$

где b_F определено в (2.23). Очевидно, что при $X_F = X$ из (2.24) получаем (2.21).

Итак, в силу (2.22), (2.24) величины $J_F^b(s)$, $J_F^v(s)$ как функции s характеризуются следующими свойствами: структурная составляющая в общем случае немонотонно убывает от значения $J_F^b(0) = y_F^o y_F^o$ до нуля, а шумовая монотонно возрастает. Следовательно, функция $J_F(s) = J_F^b(s) + J_F^v(s)$ обладает минимумом (глобальным), который находится внутри интервала изменения дискретного аргумента $s \in [1, m]$. Более того, на самом деле этот минимум лежит в более узком интервале $[1, s_0]$, $s_0 \leq m$.

Суммируем полученные выше результаты.

Утверждение 2.1. Величина дисперсии ошибки прогнозирования $J_F(s)$ как функция сложности модели s является суммой двух составляющих - структурной $J_F^b(s)$ и шумовой $J_F^v(s)$ и характеризуется следующими свойствами:

а) $J_F^b(s)$ согласно (2.22) немонотонно убывает от значения $J_F^b(0) = y_F^o y_F^o$ до нуля, причем $J_F^b(s | s \geq s_0) = 0$;

б) $J_F^v(s)$ согласно (2.24) монотонно возрастает от $J_F^v(0) = 0$ до величины $J_F^v(m) = \text{tr}(X^T X)^{-1} X_F^T X_F$;

в) минимум $J_F(s)$ достигается внутри интервала $1 \leq s \leq s_0$, $s_0 \leq m$, причем в общем случае функция $J_F(s)$ многоэкстремальна.

Величина дисперсии ошибки восстановления $J(s)$ имеет монотонно убывающую структурную $J^b(s)$ (2.12) и линейно возрастающую шумовую $J^v(s)$ (2.13) составляющие.

Непрерывная природа σ^2 и ее линейное вхождение в формулы для $J_F(s)$, $J(s)$ позволяют построить удобный аппарат для прямой "идентификации" значений s_F^o, s^o , опираясь на условие экстремума функции дискретного аргумента.

Определение 2.1. Критической дисперсией [18] $\sigma_{Fkp}^2(s_1, s_2)$ для функционала $J_F(s)$ называется значение дисперсии σ^2 , определяемое для сложностей s_1 и s_2 как неотрицательное решение относительно σ^2 уравнения

$$J_F(s_1, \sigma^2) = J_F(s_2, \sigma^2), \quad 1 \leq s_1 < s_2 \leq m. \quad (2.25)$$

Для $J(s)$ критическая дисперсия $\sigma_{kp}^2(s_1, s_2)$ определяется из уравнения

$$J(s_1, \sigma^2) = J(s_2, \sigma^2), \quad 1 \leq s_1 < s_2 \leq m. \quad (2.26)$$

В случае отрицательного решения (2.25) принимаем, что $\sigma_{Fkp}^2(s_1, s_2) = 0$.

Исключение из рассмотрения отрицательных решений уравнения (2.25) оправдано тем, что $\sigma_{Fkp}^2(\cdot)$ имеет смысл дисперсии, которая неотрицательна по определению. Важно подчеркнуть, что величина критической дисперсии *не зависит* от уровня шума σ^2 и является объективной характеристикой самого моделируемого объекта, точнее, информации о нем в виде матриц X, X_F и точных векторов $\overset{o}{y}, \overset{o}{y}_F$.

Введем теперь понятие относительной эффективности двух структур сложности s_1 и s_2 с учетом того, что $J_F(s), J(s)$ являются дисперсиями ошибок модели сложности s .

Определение 2.2. Структура s_2 (подмножество регрессоров X_{s_2}) *эффективнее* структуры s_1 (подмножества X_{s_1}), где $s_1 < s_2$, в смысле $J_F(s_2) < J_F(s_1)$ ($J(s) < J(s)$), если $\sigma^2 < \sigma_{Fkp}^2(s_1, s_2)$ ($\sigma^2 < \sigma_{kp}^2(s_1, s_2)$), и наоборот - структура s_1 эффективнее s_2 в случае $\sigma^2 \geq \sigma_{Fkp}^2(s_1, s_2)$ ($\sigma^2 \geq \sigma_{kp}^2(s_1, s_2)$).

При этом естественной *дискриминирующей функцией* для моделей сложности s_1 и s_2 по величине дисперсии ошибки прогнозирования сигнала является разность

$$\delta_F(s_1, s_2) = \sigma_{Fkp}^2(s_1, s_2) - \sigma^2, \quad (2.27)$$

поскольку при $\delta_F(s_1, s_2) > 0$ предпочтение отдается s_2 , а при $\delta_F(s_1, s_2) \leq 0$ - s_1 . Аналогично определяется $\delta(s_1, s_2)$ для задачи восстановления сигнала:

$$\delta(s_1, s_2) = \sigma_{kp}^2(s_1, s_2) - \sigma^2. \quad (2.28)$$

Следовательно, значение критической дисперсии определяет критический *порог* (меру) *различимости* двух структур сложности s_1 и s_2 с помощью указанных дискриминирующих функции, что и объясняет принятое название величин $\sigma_{Fkp}^2(\cdot), \sigma_{kp}^2(\cdot)$.

Приведем также интерпретацию критических дисперсий в терминах теории связи.

Определение 2.3. Если $\sigma_{Fkp}^2(s_1, s_2) > 0$, то структура s_2 называется *помехоустойчивой* по отношению к s_1 в смысле существования некоторого диапазона изменения дисперсии шума $0 < \sigma^2 < \sigma_{Fkp}^2(\cdot)$, в котором меньшую дисперсию ошибки имеет структура s_2 .

Следовательно, критическая дисперсия является теоретической *мерой относительной помехоустойчивости* структур s_1 и s_2 : чем она больше, тем больше диапазон изменения уровня шума, не меняющего предпочтения структуры s_2 относительно s_1 .

Рассмотрим теперь применимость критических дисперсий в качестве средства для аналитического исследования проблем структурной идентификации моделей.

Определение 2.4. Совокупность операций по вычислению критических дисперсий, анализу их свойств и интерпретации полученных результатов в задаче выбора структур моделей с минимальной дисперсией ошибки называется *методом критических дисперсий*.

Этот метод лежит в основе полученных в отчете теоретических результатов. Приведем общую характеристику связи основных свойств величины критической дисперсии с поведением оптимальных значений s_F^o, s^o .

В задаче минимизации дисперсии ошибки прогнозирования важны локальные свойства функций $J_F(s), J(s)$. А эти свойства определяются первыми разностями этих функций, то есть необходимо вычислять и анализировать критические дисперсии для значений $s_1 = s$, $s_2 = s+1$, и при этом уравнения (2.25), (2.26) дают необходимые условия локального экстремума. Исходя из этого, на основе утверждения 2.1 и определения 2.1 сформулируем результат, характеризующий применимость метода критических дисперсий в задаче оптимизации структуры модели (на примере $J_F(s)$).

Утверждение 2.2. Связь между $\sigma_{Fkp}^2(s, s+1)$ и экстремальными свойствами $J_F(s)$ характеризуется следующими особенностями:

а) структурная составляющая $J_F^b(s)$ строго *монотонно* убывает только тогда, когда для всех $s = \overline{1, m-1}$ имеем $\sigma_{Fkp}^2(\cdot) > 0$;

б) $J_F(s)$ является функцией s с *единственным* минимумом только тогда, когда все значения $\sigma_{Fkp}^2(\cdot)$ строго упорядочены по убыванию:

$$\sigma_{Fkp}^2(1,2) > \sigma_{Fkp}^2(2,3) > \dots > \sigma_{Fkp}^2(m-1); \quad (2.29)$$

в) если выполняется (2.29), то минимум $J_F^o = J_F(s_F^o)$ при заданном σ^2 соответствует такой сложности $s = s_F^o$, для которой выполняется

$$\sigma_{Fkp}^2(s-1, s) > \sigma^2 \geq \sigma_{Fkp}^2(s, s+1); \quad (2.30)$$

г) если (2.29) не выполняется, то (2.30) является условием локального минимума, и для определения s_F^o необходимо найти минимальное значение функции $J_F(s)$ среди всех точек, для которых выполняется (2.30).

Доказательство этих утверждений, справедливых также для функции $J(s)$, следует непосредственно из определения 2.1 с учетом того, что достаточным условием минимума в точке s является положительность второй разности

$$\Delta^2 J_F(s) = \Delta J_F(s-1) - \Delta J_F(s) = J_F(s-1) - 2J_F(s) + J_F(s+1),$$

причем из сравнения (2.25) и (2.27) очевидно, что для любого s

$$\text{sign } \Delta J_F(s) = \text{sign } \delta_F(s, s+1),$$

так что $\Delta^2 J_F(s) > 0$ при $\delta_F(s-1, s) > 0$, $\delta_F(s, s+1) < 0$, что совпадает с (2.30).

С учетом полученного предлагается следующая *методика анализа* закономерностей выбора оптимальных структур моделей в условиях неопределенности:

- 1) цель анализа - ответить на вопрос: как выбирались бы оптимальные структуры в идеальном случае, когда неизвестные на практике величины σ^2, y, y_F известны, но оценки $\theta_{s,s} = \overline{1, m}$ вычисляются по зашумленным данным;
- 2) исследуется прежде всего базовый случай $X = \overset{o}{X}, m = s_o$;
- 3) исследуются закономерности изменения оптимальных значений s_F^o, s^o в зависимости от дисперсии (уровня) шума σ^2 с помощью вычисления $\sigma_{Fkp}^2(s, s+1), \sigma_{kp}^2(s, s+1)$;
- 4) исследуются асимптотические свойства оптимальных структур;
- 5) на основе полученных результатов формулируются требования к свойствам вычислимых (физически реализуемых) критериев, применяемых на практике в качестве оценок дисперсии ошибки модели.

Эта методика применяется ниже в теоретических исследованиях.

2.3. Закономерности выбора оптимальных структур моделей в зависимости от уровня шума

2.3.1. Случай усреднения по всем возможным реализациям шума

Как показано выше, значения s^o, s_F^o определяются соотношением s и величин критических дисперсий для $s_1 = s$, $s_2 = s+1$.

Лемма 2.1. Критическая дисперсия $\sigma_{kp}^2(s, s+1)$ для задачи восстановления сигнала существует для всякого s и равна

$$\sigma_{kp}^2(s, s+1) = \overline{\omega}_{s+1}^2 \beta_{s+1}. \quad (2.31)$$

Доказательство следует из определения 2.1 и формул (2.20), (2.21), поскольку

$$J(s+1) = J(s) - \overline{\omega}_{s+1}^2 \beta_{s+1} + \sigma^2. \quad (2.32)$$

Лемма 2.2. Критическая дисперсия $\sigma_{Fkp}^2(s, s+1)$ для задачи прогнозирования сигнала определяется выражением

$$\sigma_{Fkp}^2(s, s+1) = 2\overline{\omega}_{s+1}^2 \beta_{s+1} \frac{a_F^T b_F}{b_F^T b_F} - \overline{\omega}_{s+1}^2 \beta_{s+1}, \quad (2.33)$$

если выполняется условие (2.24), и равна нулю в противном случае.

Доказательство следует из определения 2.1 и формул (2.22), (2.24), поскольку

$$J_F(s+1) = J_F(s) - 2\overline{\omega}_{s+1} a_F^T b_F - \overline{\omega}_{s+1}^2 b_F^T b_F + \sigma^2 \frac{b_F^T b_F}{\beta_{s+1}}. \quad (2.34)$$

Исследуем поведение значений s^o, s_F^o при повышении дисперсии (уровня) шума σ^2 .

Утверждение 2.3. Если выполняется условие строгого упорядочения (2.29), то:

а) истинная структура, содержащая все множество регрессоров $X=X_m$, является оптимальной как для восстановления ($s^o = m$), так и для прогнозирования ($s_F^o = m$) истинного сигнала, если σ^2 удовлетворяет неравенству

$$0 \leq \sigma^2 < \sigma_{kp}^2(m-1, m); \quad (2.35)$$

б) при постепенном увеличении σ^2 выше уровня $\sigma_{kp}^2(m-1, m)$ оптимальной становится поочередно все более простая структура: $s_F^o = m-1, m-2, \dots$, причем $s = s_F^o$ при выполнении (2.30), а в случае $\sigma^2 > \sigma_{Fkp}^2(1, 2)$ имеем $s_F^o = 1$; то же относится и к значению s^o , зависящему от соотношения σ^2 и $\sigma_{kp}^2(s, s+1)$;

в) оптимальные структуры для задач восстановления и прогнозирования, вообще говоря, не совпадают, $s^o \neq s_F^o$.

Часть а) утверждения следует из того, что величины (2.31) и (2.33) совпадают при $s=m-1$, часть б) является следствием утверждения 2.2, а в) очевидно ввиду того, что, вообще говоря, (2.31) и (2.33) не равны. Качественно закономерность упрощения оптимальных структур при увеличении σ^2 имеет вид рис.2.1.

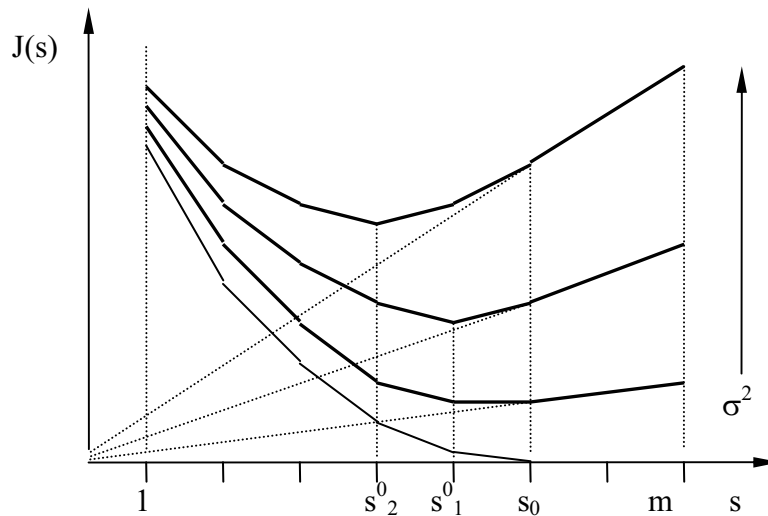


Рисунок 2.1 - Упрощение оптимальной структуры при возрастании уровня шума σ^2

2.3.2. Случай отдельной реализации шума

Полученные выше результаты, характеризующие теоретические особенности выбора оптимальной структуры модели объекта, справедливы “в среднем” по всем возможным реализациям вектора шума. Однако на практике большинство задач решается по ограниченной выборке “без повторений”, когда данные содержат единственную реализацию вектора шума, и анализ такого случая представляет отдельный интерес.

Определим аналогично (2.15) “идеальный” критерий для случая единственной реализации вектора шума как абсолютное значение теоретической ошибки модели:

$$R(s) = \left\| \begin{matrix} o \\ y - X_s \hat{\theta}_s \end{matrix} \right\|^2, \quad (2.36)$$

т.е. $E[R(s)] = J(s)$. Допустим для наглядности, что $\xi = \tilde{\sigma}\varepsilon$, где ε - произвольная реализация $n+1$ вектора шума с единичной дисперсией, тогда $\tilde{\sigma}$ имеет смысл амплитуды (уровня) шума ξ . Вектор ошибки в (2.36) можно представить как

$$\begin{matrix} o \\ y - X_s \hat{\theta}_s \end{matrix} = \begin{matrix} o \\ y - X_s \bar{\theta}_s - X_s (X_s^T X)^{-1} X_s^T \xi \end{matrix} = D_s \begin{matrix} o \\ y - P_s \xi \end{matrix},$$

тогда ввиду $D_s = I_n - P_s$, где $P_s = X_s (X_s^T X)^{-1} X_s^T$, получаем:

$$R(s) = R^b(s) + R^v(s) = \left\| \begin{matrix} o \\ y - X_s \bar{\theta}_s \end{matrix} \right\|^2 + \tilde{\sigma}^2 \varepsilon^T P_s \varepsilon, \quad (2.37)$$

поскольку D_s, P_s - идемпотентные матрицы, а произведение $\begin{matrix} o \\ y^T \end{matrix} D_s P_s \xi = 0$. Поскольку $R^b(s) = J^b(s)$, для изучения особенностей выбора модели оптимальной сложности

$$\tilde{s}^o = \arg \min_{s=1, m} R(s), \quad (2.38)$$

следует изучить $R^v(s)$ как функцию s . Запишем $\varepsilon^T P_s \varepsilon$ в виде $\varepsilon^T P_s \varepsilon = \hat{a}_{\varepsilon}^T X_s^T X_s \hat{a}_{\varepsilon}$, где "оценка" \hat{a}_{ε}^T аналогична оценке МНК: $\hat{a}_{\varepsilon}^T = (X_s^T X_s)^{-1} X_s^T \xi$. Вычислим теперь $\varepsilon^T P_{s+1} \varepsilon$ с учетом того, что, по аналогии с рекуррентным МНК, $\hat{a}_{\varepsilon, s+1} = (\hat{a}_{\varepsilon}^T \hat{\omega}_{\varepsilon})^T$, где

$$\hat{\omega}_{\varepsilon} = x^T D_s \varepsilon / x^T D_s x. \quad (2.39)$$

Тогда получаем, что $\varepsilon^T P_{s+1} \varepsilon = \varepsilon^T P_s \varepsilon + \hat{\omega}_{\varepsilon}^2 \beta_{s+1}$, где $\beta_{s+1} = x^T D_s x > 0$, поэтому

$$R^v(s+1) = R^v(s) + \tilde{\sigma}^2 \hat{\omega}_{\varepsilon}^2 \beta_{s+1}, \quad (2.40)$$

т.е. шумовая составляющая критерия $R(s)$ является строго монотонно возрастающей функцией s . Следовательно, $R(s)$ при всяком $\tilde{\sigma}^2$ имеет минимум, соответствующий

некоторой сложности \tilde{s}^o , равной такому s , при котором удовлетворяется неравенство $\tilde{\sigma}_{kp}^2(s-1, s) > \tilde{\sigma}^2 \geq \tilde{\sigma}_{kp}^2(s, s+1)$, где критическая дисперсия вычисляется из следующего линейного относительно $\tilde{\sigma}^2$ уравнения:

$$R(s, \tilde{\sigma}^2) = R(s+1, \tilde{\sigma}^2), \quad 1 \leq s \leq m-1. \quad (2.41)$$

Критическая дисперсия является мерой различимости моделей сложности s и $s+1$ и в данном случае равна:

$$\tilde{\sigma}_{kp}^2(s, s+1) = \frac{\bar{\omega}_{s+1}^2 \beta_{s+1}}{\hat{\omega}_{\varepsilon}^2 \beta_{s+1}} = \frac{(x^T D_s y)^2}{(x^T D_s \varepsilon)^2}, \quad (2.42)$$

где $\bar{\omega}_{s+1}^2 = (x^T D_s y)^2 / \beta_{s+1}^2$. Положительность величины $\tilde{\sigma}_{kp}^2(s, s+1)$ для любого s означает, что критерий $R(s)$ в качественном смысле имеет те же свойства, что и $J(s)$, хотя и относится к отдельно взятой реализации вектора шума.

При этом важно, что

$$E[\tilde{\sigma}_{kp}^2(s, s+1)] = \sigma_{kp}^2(s, s+1),$$

где $\sigma_{kp}^2(s, s+1)$ - величина критической дисперсии для функционала $J(s)$, поскольку $E[\hat{\omega}_\varepsilon^2] = \text{tr}(D_s x x^T D_s) / \beta_{s+1}^2 \equiv 1 / \beta_{s+1}$.

Изучение асимптотического поведения ошибки $R(s)$ требует, ввиду случайности вектора ε , проверки сходимости в вероятностном смысле. В частности, ввиду стохастической независимости X_s и ε имеет место сходимость по вероятности $p\lim \frac{1}{n} X_{ns}^T \varepsilon_n = 0_s$, поэтому $p\lim \frac{1}{n} R^v(n, s) = \sigma^2 p\lim \frac{1}{n} \varepsilon_n^T P_{ns} \varepsilon_n = 0$, следовательно,

$$p\lim \frac{1}{n} R(n, s) = \tilde{J}(s).$$

где $\tilde{J}(s) = \lim_{n \rightarrow \infty} \frac{1}{n} J(n, s)$. Таким образом, величина $\frac{1}{n} R(n, s)$ при каждом s сходится по вероятности к детерминированной функции $\tilde{J}(s)$ (этот предел, как показано в [12], существует ввиду предположения П4), являющейся предельным значением для усредненной дисперсии ошибки (2.11). Следовательно, оптимальная величина \tilde{s}^o является состоятельной оценкой s^o в том смысле, что

$$p\lim \frac{1}{n} R(\tilde{s}^o) = \lim_{n \rightarrow \infty} \frac{1}{n} J(s^o).$$

Итак, абсолютная величина теоретической ошибки модели в условиях отдельной реализации вектора шума имеет такие же качественные свойства, что и дисперсия ошибки (т.е. величина потерь, усредненная по всевозможным реализациям вектора шума).

В заключение выделим основные выводы: 1) любая смещенная (упрощенная) структура, включающая $s < m$ регрессоров, может при достаточно большом σ^2 иметь меньшую дисперсию ошибки модели, чем несмещенная (истинная) структура; 2) чем больше уровень шума σ^2 , тем проще модель с минимальной дисперсией ошибки. Эти выводы хорошо согласуются с известной теорией Шеннона [19] о канале с шумом: при увеличении уровня шума точность передачи сигнала (аналог - дисперсия ошибки модели) можно сохранить, уменьшив пропускную способность (аналог - сложность модели) канала связи. На качественном уровне конструктивность такой аналогии впервые отмечена в [15]. При этом модель оптимальной сложности s^o или s_F^o является наиболее помехоустойчивой (в терминах [16] - потенциально помехоустойчивой). В связи с этим свойство *упрощения* оптимальной структуры при возрастании σ^2 естественно назвать *свойством помехоустойчивости*, которое объективно присуще задаче выбора модели с минимальной дисперсией ошибки. Всякий критерий сравнения моделей, применяемый на практике в качестве оценки дисперсии ошибки, с должен обладать этим свойством.

Следует подчеркнуть стохастическую природу свойства помехоустойчивости: параметры моделей оцениваются по зашумленной выборке, и эти оценки для точной структуры являются *несмещенными*, однако чем больше параметров оценивается, тем быстрее возрастает дисперсия этих оценок при увеличении уровня шума, и поэтому

смещенная, или "загрубленная" (упрощенная) модель может быть эффективнее полной, т.е. иметь меньшую среднюю ошибку по всевозможным реализациям вектора шума.

2.4. МГУА как эффективный метод построения моделей с минимальной дисперсией ошибки

В данном подразделе разрабатывается методика аналитического решения задачи оценки эффективности конкретного критерия $CR(s)$ с точки зрения минимизации дисперсии ошибки прогнозирования и излагаются результаты применения этой методики к анализу эффективности внешних критериев МГУА. При этом применяется рассмотренный выше метод критических дисперсий как инструмент исследования проблем структурной идентификации прогнозирующих моделей. Ниже будем считать справедливыми базовые предположения П1-П4, принятые в подразделе 2.1.

Методика анализа критериев выбора моделей с применением метода критических дисперсий. В соответствии с поставленной нами целью структурной идентификации моделей с минимальной дисперсией ошибки прогнозирования, для всякого реально применяемого критерия $CR(s)$ следует проверить, насколько выбираемые по этому критерию модели близки к моделям, оптимальным в смысле минимума $J_F(s)$. Поэтому эффективность заданного критерия $CR(s)$ будем анализировать с точки зрения того, как случайная величина \hat{s}_{CR} , определенная в виде

$$\hat{s}_{CR} = \arg \min_{s=1,m} CR(s), \quad (2.43)$$

оценивает неизвестное значение s_F^o , соответствующее минимуму теоретического критерия $J_F(s)$ (2.12). Введем величину

$$s_{CR}^* = \arg \min_{s=1,m} E [CR(s)] \quad (2.44)$$

как усредненное значение \hat{s}_{CR} и уточним понятие эффективности критерия.

Определение 2.5. Критерий $CR(s)$ при произвольных конечных σ^2 и n называется *оптимальным* для оценивания прогнозирующих моделей, если $s_{CR}^* = s_F^o$ (несмещенное оценивание s_F^o), *адекватным* при $s_{CR}^* \leq s_F^o$ (помехоустойчивое оценивание s_F^o) и *неадекватным* при $s_{CR}^* > s_F^o$.

Отметим, что в случае $s_{CR}^* > s_F^o$ критерий следует считать неадекватным, поскольку s_F^o по определению критерия $J_F(s)$ является теоретически максимальным (предельно достижимым, или потенциально помехоустойчивым) значением.

Учитывая определение 2.5, а также результаты данного раздела, эффективность заданного критерия целесообразно анализировать в такой последовательности: найти его математическое ожидание; проверить компромиссный характер структурной и шумовой составляющих полученного выражения; определить условия оптимальности или адекватности критерия. Последнее задание решается формализованно с помощью метода критических дисперсий, применявшегося для анализа функционалов $J_F(s)$, $J(s)$, $R(s)$.

Определение 2.6. Величина критической дисперсии $\sigma_{CR}^2(s, s+1)$ как теоретическая мера различимости моделей сложности s и $s+1$ по критерию $CR(s)$ равна значению σ^2 , являющемуся неотрицательным решением уравнения

$$E[CR(s, \sigma^2)] = E[CR(s+1, \sigma^2)] . \quad (2.45)$$

Поскольку это условие локального минимума $CR(s)$ как функции s , достаточно очевидны следующие свойства критической дисперсии: а) если положительное решение (2.45) отсутствует для всех $s = \overline{1, m-1}$, то $CR(s)$ не имеет экстремума и неприменим в качестве критерия выбора модели; б) чем больше $\sigma_{CR}^2(s, s+1)$, тем больше различимость двух моделей сложности s и $s+1$, т.е. больше диапазон изменения σ^2 , в котором "в среднем" лучшей по данному критерию является модель $s+1$; в) модель сложности s оптимальна в смысле $s_{CR}^* = s$ при выполнении условия $\sigma_{CR}^2(s-1, s) > \sigma^2 \geq \sigma_{CR}^2(s, s+1)$.

Аналогичные свойства имеет критическая дисперсия $\sigma_{Fkp}^2(s, s+1)$, что и позволяет анализировать эффективность заданного $CR(s)$ в смысле минимизации дисперсии ошибки прогнозирования, основываясь на следующем результате.

Утверждение 2.4. Критерий $CR(s)$ является оптимальным или адекватным, если равномерно по s выполняются условия $\sigma_{CR}^2(s, s+1) = \sigma_{Fkp}^2(s, s+1)$ или $\sigma_{CR}^2(s, s+1) \leq \sigma_{Fkp}^2(s, s+1)$, соответственно.

Доказательство с учетом определений 2.5 и 2.6 следует из того, что критические дисперсии вычисляются из условия равенства нулю первых разностей дискретных по s функций: при $\sigma_{CR}^2(\cdot) = \sigma_{Fkp}^2(\cdot)$ функции $E[CR(s)]$ и $J_F(s)$ отличаются лишь на константу, т.е. $s_{CR}^* = s_F^o$, а при $\sigma_{CR}^2(\cdot) \leq \sigma_{Fkp}^2(\cdot)$ всегда имеем $s_{CR}^* \leq s_F^o$. Утверждение 2.4 дает необходимые и достаточные условия оптимальности или адекватности.

Поскольку при вычислении $\sigma_{CR}^2(s, s+1)$ используется матрица X , а для $\sigma_{Fkp}^2(\cdot)$ как X , так и X_F , очевидно, что в общем случае, когда X и X_F произвольные, условия утверждения 2.4 трудно проверить аналитически. Поэтому ниже подробно рассмотрен случай, связанный с планированием эксперимента, когда полная матрица плана X является квадратично зависимой от X_F [20], т.е. удовлетворяет условию:

$$\rho^2 X^T X = X_F^T X_F, \quad \rho^2 \neq 0 . \quad (2.46)$$

При этом независимо от ρ^2 справедливо равенство $\sigma_{Fkp}^2(s, s+1) = \sigma_{kp}^2(s, s+1)$, где $\sigma_{kp}^2(\cdot)$ - критическая дисперсия для функционала $J(s)$, что позволяет аналитически решать задачу оценки эффективности критериев.

Анализ эффективности критериев МГУА. Ниже рассмотрено применение изложенной методики для анализа эффективности одного из критериев МГУА, а именно

наиболее часто применяемого критерия регулярности $AR(s)$ (1.29), являющегося критерием с неявным учетом сложности модели. Очевидно, что при выполнении (2.46) критерии сравниваются фактически с величиной дисперсии ошибки восстановления сигнала $J(s)$, имеющей вид (2.17) и содержащей структурную $J^b(s)$ и шумовую $J^v(s)$ составляющие.

По аналогии с (2.16), несложно вычислить математическое ожидание критерия $AR(s)$:

$$\overline{AR}(s) = E[AR(s)] = J_B(s) + \sigma^2 n_B, \quad (2.47)$$

причем здесь $J_B(s)$ равно дисперсии ошибки прогнозирования на подвыборке B , а параметры модели оцениваются на A , или, в силу очевидной аналогии с (2.16),

$$J_B(s) = J_B^b(s) + J_B^v(s) = \left\| y_B - X_{Bs} \bar{\theta}_{As} \right\|^2 + \sigma^2 \text{tr}(X_{As}^T X_{As})^{-1} X_{Bs}^T X_{Bs}, \quad (2.48)$$

где $\bar{\theta}_{As} = E[\hat{\theta}_{As}]$, а $\hat{\theta}_{As}$ определено в (1.26). При этом $J_B^b(s)$ является в общем случае немонотонно убывающей (до $J_B^b(m)=0$), а $J_B^v(s)$ - монотонно возрастающей (как и $J_F^v(s)$ в (2.16)) функциями s . В рассматриваемом же нами случае планируемого эксперимента имеется возможность при построении плана X , удовлетворяющего (2.46) при заданном X_F , получить одновременно разбиение матрицы X на такие X_A и X_B , что при условии $\text{rank } X_B = m$ выполняется аналогичное (2.46) соотношение:

$$\rho_B^2 X_A^T X_A = X_B^T X_B, \quad \rho_B^2 \neq 0. \quad (2.49)$$

Определение 2.7. Квадратично зависимым, или ρ_B^2 -пропорциональным, называется разбиение матрицы плана X на такие подматрицы X_A и X_B , $\text{rank } X_A = \text{rank } X_B = m$, что при некотором $\rho_B^2 \neq 0$ выполняется (2.49).

При условии (2.49) ввиду (2.48) имеем $J_B(s, \rho_B^2) = \rho_B^2 J_A(s)$, где

$$J_A(s) = E \left\| y_A - X_{As} \hat{\theta}_{As} \right\|^2 = \left\| y_A - X_{As} \bar{\theta}_{As} \right\|^2 + \sigma^2 s, \quad (2.50)$$

так что $J_B(s, \rho_B^2) = \rho_B^2 J_A(s)$ является монотонно убывающей функцией s , т.е. минимумы $J_B(s, \rho_B^2)$ и $J_A(s)$ соответствуют одной и той же точке $s_B^0 = s_A^0$ независимо от ρ_B^2 . Возможность одновременного удовлетворения (2.46) и (2.49) следует из того, что

$$X^T X = X_A^T X_A + X_B^T X_B = (1 + \rho_B^2) X_A^T X_A. \quad (2.51)$$

Тогда, с учетом (2.46), получаем [20]:

$$X_A^T X_A = \frac{1}{\rho^2(1 + \rho_B^2)} X_F^T X_F, \quad (2.52)$$

$$X_B^T X_B = \frac{\rho_B^2}{\rho^2(1 + \rho_B^2)} X_F^T X_F, \quad (2.53)$$

определяя тем самым и ρ^2 -пропорциональный план X , и ρ_B^2 -пропорциональное разбиение матрицы X .

Отметим, что из (2.47) - (2.48) сразу следует, что критерий $AR(s)$ обладает свойством помехоустойчивости (смещения минимума при увеличении дисперсии шума σ^2) как

объективно необходимым свойством эффективного критерия структурной идентификации. Теперь, согласно избранной методике, нам следует проверить оптимальность или адекватность исследуемого критерия регулярности (2.47), основанного на разбиении выборки (с неявным ограничением сложности).

Лемма 2.3. Критическая дисперсия для критерия $AR(s)$ равна

$$\sigma_{AR}^2(s, s+1) = 2\bar{\omega}_{A,s+1}\beta_{A,s+1} \frac{a_B^T b_B}{b_B^T b_B} - \bar{\omega}_{A,s+1}^2 \beta_{A,s+1}, \quad (2.54)$$

где $a_B = y - X_{Bs} \bar{\theta}_{As}$, $b_B = x_B - X_{Bs} (X_{As}^T X_{As})^{-1} X_{As}^T x_A$, а величины $\bar{\omega}_{A,s+1}$, $\beta_{A,s+1}$ вычисляются так же, как и $\bar{\omega}_{s+1}$, β_{s+1} , но на подвыборке A .

Лемма доказывается совершенно аналогично лемме 2.2 ввиду очевидного соответствия функционалов $J_F(s)$ и $J_B(s)$ при замене множеств W , F на множества A , B . Имеет место также следующий частный случай леммы 2.3.

Следствие 1. При выполнении разбиения (2.49) из (2.54) независимо от ρ_B^2 получаем:

$$\sigma_{AR}^2(s, s+1) \equiv \sigma_{Bkp}^2(s, s+1) = \bar{\omega}_{A,s+1}^2 \beta_{A,s+1}. \quad (2.55)$$

Для доказательства достаточно подставить (2.49) в (2.54) и получить равенство $a_B^T b_B / b_B^T b_B = \bar{\omega}_{A,s+1}$.

Еще один важный частный случай леммы 3.3, доказательство которого очевидно.

Следствие 2. При $s = m-1$ для любого X_B независимо от разбиения выражение (2.54) принимает вид

$$\sigma_{AR}^2(m-1, m) = \sigma_{Akp}^2(m-1, m) = (\theta_o^m)^2 \beta_{Am}, \quad (2.56)$$

где θ_o^m - m -я компонента истинного вектора параметров θ_0 .

Теперь можно оценить эффективность критерия регулярности.

Утверждение 2.5. Критерий $AR(s)$ является адекватным задаче оценивания оптимальной сложности s_F^o в условиях выполнения (2.49).

Для доказательства заметим, что в случае разбиения вида (2.49) из (2.47) получаем:

$$\overline{AR}(s, \rho_B^2) = \rho_B^2 J_A(s) + \sigma^2 n_B. \quad (2.57)$$

То есть при любом s минимум критерия регулярности "в среднем" соответствует оптимальной структуре в смысле минимума одновременно как $J_B(s)$, так и $J_A(s)$:

$$s_{AR}^* = \arg \min_{s=1,m} \overline{AR}(s, \rho_B^2) = \arg \min_{s=1,m} J_B(s, \rho_B^2) = \arg \min_{s=1,m} J_A(s). \quad (2.58)$$

При этом значение критической дисперсии для структуры сложности s принимает вид (2.55), где $\bar{\omega}_{A,s+1} = \alpha_{A,s+1} / \beta_{A,s+1}$.

Поскольку при вычислении оптимального значения s_{AR}^* по критерию регулярности используется только часть A всей выборки, а минимальная дисперсия прогноза в точках X_F определяется значением s_F^o , которое получается с учетом всех точек выборки,

необходимо установить соответствие между s_{AR}^* и $s_F^o = s^o$. Для этого следует применить утверждение 2.4, что несложно сделать ввиду

$$\alpha_{s+1} = (1 + \rho_B^2) \alpha_{A,s+1}, \quad \beta_{s+1} = (1 + \rho_B^2) \beta_{A,s+1},$$

так что $\bar{\omega}_{A,s+1} = \bar{\omega}_{B,s+1} = \bar{\omega}_{s+1}$, и с учетом (2.55) имеем:

$$\sigma_{AR}^2(s, s+1) = \sigma_{Akp}^2(s, s+1) = \frac{\bar{\omega}_{s+1}^2 \beta_{s+1}}{1 + \rho_B^2} = \frac{1}{1 + \rho_B^2} \sigma_{kp}^2(s, s+1). \quad (2.59)$$

Следовательно, имеет место прямое соотношение между критическими дисперсиями на всей выборке и на ее части A , причем $\sigma_{AR}^2(s, s+1) \leq \sigma_{kp}^2(s, s+1)$, из чего следует, что при любом σ^2 справедливо неравенство $s_{AR}^* \leq s^o$, т.е. критерий регулярности в среднем всегда выбирает более простую модель, чем оптимальная (что естественно ввиду упомянутой разницы в числе точек), что и завершает доказательство утверждения 2.5.

Таким образом, в условиях планируемого эксперимента (при выполнении условий (2.46) для плана X и (2.49) для разбиения X на X_A и X_B) критерий регулярности $AR(s)$ в силу утверждения 2.5 гарантированно является адекватным задаче оценивания оптимальных прогнозирующих моделей в смысле $s_{AR}^* \leq s_F^o$. В общем случае произвольного соотношения X и X_F в условиях конечного числа наблюдений применение $AR(s)$ по-прежнему оправдано тем, что он по своей природе ориентирован на выбор прогнозирующих моделей, причем ввиду (2.47) он является J_B -оптимальным.

Аналогичным образом можно проанализировать другие внешние критерии, применяемые в МГУА, причем такая задача облегчается тем, что все они выражаются через величины типа (1.27), т.е. для вычисления $\bar{CR}(s)$, и соответственно $\sigma_{CR}^2(\cdot)$, достаточно использовать выражения, аналогичные уже полученным в (2.47) для $AR(s)$ (т.е. $\bar{\Delta}(B|A) = E[\Delta(B|A)]$). Например, для симметричного критерия регулярности $AD(s)$ с учетом (1.30) при условии (2.49) получаем:

$$\sigma_{AD}^2(s, s+1) = \frac{1}{\rho_B^2 + \frac{1}{\rho_B^2}} \bar{\omega}_{s+1}^2 \beta_{s+1}, \quad (2.50)$$

что говорит об адекватности $AD(s)$, так как $\sigma_{AD}^2(\cdot) \leq \frac{1}{2} \sigma_{kp}^2(\cdot)$ ввиду $(\rho_B^2 + 1 / \rho_B^2) \geq 2$.

Что касается критериев группы согласованности (1.32) и (1.33), то в условиях разбиения (2.49) они неработоспособны, поскольку при этом, как показано выше, $\bar{\omega}_{A,s+1} = \bar{\omega}_{B,s+1} = \bar{\omega}_{s+1}$, вследствие чего структурные составляющие $CB^b(s) = CB^v(s) = 0$, и соответственно $\sigma_{CB}^2(\cdot) = \sigma_{CV}^2(\cdot) = 0$. Поэтому очевидной рекомендацией для их эффективного использования является выбор разбиения по условию максимального рассогласования выражения (2.49), дающему наилучшую различимость моделей [21]. В случае, когда (2.49) не выполняется, критерии согласованности являются адекватными в

смысле определения 2.5, что для $CB(s)$ очевидно из (1.32) с учетом (2.50), а для $CV(s)$ следует из выражения

$$\sigma_{CV}^2(s, s+1) = \bar{\omega}_{s+1}^2 \beta_{s+1} - (\bar{\omega}_{A,s+1}^2 \beta_{A,s+1} + \bar{\omega}_{B,s+1}^2 \beta_{B,s+1}), \quad (2.51)$$

которое нетрудно получить ввиду

$$\overline{CV}(s) = J^b(s) - (J_A^b(s) + J_B^b(s) + \sigma^2 s).$$

Совершенно аналогично можно применить метод критических дисперсий для аналитической проверки эффективности применения любого другого критерия качества моделей $CR(s)$ в качестве вычислимой оценки теоретического функционала $J_F(s)$ для задачи структурной идентификации прогнозирующих (J_F -оптимальных) моделей.

Полученные выше результаты доказывают, что внешние критерии МГУА являются адекватными задаче построения моделей с минимальной дисперсией ошибки прогноза. Это говорит об эффективности применения алгоритмов МГУА в задачах прогнозирования.

2.5 Анализ проблемы разбиения выборки для алгоритмов МГУА.

Эффективность применения критериев типа “перекрестного подтверждения”, на которых базируются алгоритмы МГУА, при решении практических задач моделирования зависит, в частности, от удачного или неудачного выбора варианта разбиения выборки данных на части для формирования внешних критериев качества моделей. Проблеме разбиения посвящено довольно мало работ [22-27], что свидетельствует как о трудности ее анализа, так и о том, что при наличии некоторого опыта применения алгоритмов МГУА она на практике не является непреодолимой. Рассмотрим ее подробнее.

Из перечисленных в [2] способов разбиения выборки данных на две части в настоящее время реально применяется только способ “разбиения по дисперсии” [22], основанный на ранжировании точек наблюдений по их разбросу вокруг общей средней точки. Он удобен тем, что позволяет автоматически разделить выборку на две подвыборки до начала процедуры МГУА, хотя при этом никак не связан с целью моделирования. Рекомендация [23] сводилась к подбору разбиения, минимизирующего число рядов многорядного алгоритма (а тем самым и сложность модели), то есть требовала многократного применения алгоритма МГУА. В алгоритме [24] применялось случайное разделение.

В работах [25-27] сделаны попытки теоретического анализа проблемы разбиения. В [25] предложено наилучшее разбиение, максимизирующее чувствительность критерия несмещенности, однако оно должно строиться для каждой модели в отдельности, что трудно осуществимо на практике в алгоритмах МГУА ввиду большого объема перебора. Результат, полученный в [26], дает оптимальное разбиение, но он имеет чисто теоретическое значение, поскольку предполагает известной функцию распределения внешнего критерия, что реально не выполнимо. Отметим еще, что в [21] показана необходимость добиваться максимальной “непохожести” двух подвыборок для улучшения различимости моделей по критерию непротиворечивости, что аналогично подходу [25], а в [27] предложен очевидный вариант разбиения в случае повторного эксперимента.

Рассмотрим проблему разбиения выборки с точки зрения минимизации дисперсии ошибки прогноза по модели, соответствующей минимуму критерия регулярности.

Оптимальное разбиение выборки. Назовем *оптимальным* такое разбиение выборки W , при котором в результате идентификации по условию минимума дисперсии ошибки модели на каждой из подвыборок A и B будет получена одна и та же структура:

$$s_A^o = s_B^o, s_G^o = \arg \min_{s=1,m} J_G(s),$$

$$J_G(s) = E \left\| y_G^o - X_{Gs} \hat{\theta}_{Gs} \right\|^2, G = A, B. \quad (2.62)$$

Можно ли построить такое разбиение? В предыдущем подразделе показано, что строгое решение такой задачи предполагает планирование эксперимента. Полученные выше результаты можно суммировать в виде следующего утверждения.

Утверждение 2.6. Если разбиение выборки W на две подвыборки A и B таково, что X_A и X_B при $\text{rank} X_A = \text{rank} X_B = m$ удовлетворяют условию квадратичной зависимости (ρ_B^2 -пропорциональности) (2.49), то для критерия регулярности справедливо (2.57), и минимуму $\overline{AR}(s, \rho_B^2)$ соответствует оптимальная сложность $s_{AR}^* = s_A^o = s_B^o$ (2.58).

Замечание 1. Частные случаи ρ_B^2 -пропорционального разбиения: а) ортогональные регрессоры (см.[28,29]); б) ρ_B -пропорциональное разбиение ($\rho_B X_A = X_B$, $\rho_B \neq 0$), в том числе схема повторного эксперимента ($X_A = X_B$, см.[27]).

Замечание 2. Для моделей со свободным членом $\rho_B^2 = n_B / n_A$, поскольку первые диагональные элементы матриц $X_A^T X_A$ и $X_B^T X_B$ равны n_A и n_B соответственно.

Эти результаты характеризуют оптимальный план эксперимента и соответствующее квадратично зависимое разбиение выборки, сама же задача построения такого плана здесь не рассматривается, поскольку исследуется случай пассивной выборки. При этом подчеркнем, что условие (2.49) относится ко всем регрессорам одновременно, т.е. они справедливы для любого пути усложнения структуры модели.

Построение квазиоптимального разбиения в случае пассивной выборки. В общем случае разбиение (2.49) можно обеспечить только с помощью планирования эксперимента. В нашей же постановке задачи данные предполагаются полученными как пассивные наблюдения, поэтому можно говорить лишь о квазиоптимальном разбиении. Задачу о поиске такого разбиения можно сформулировать следующим образом: из всех возможных вариантов разбиения выборки W на подвыборки A_l, B_l такие, что $\text{rank} A_l = \text{rank} B_l = m$, выбрать наилучший вариант по условию

$$\min_{l \in [1, L]} \min_{\rho_B^2 \neq 0} \left\| \rho_B^2 X_{A_l}^T X_{A_l} - X_{B_l}^T X_{B_l} \right\|^2, \quad (2.63)$$

где L - число допустимых вариантов, а $\|\cdot\|$ - некоторая матричная норма (например, $\|C\| = \max_{i,j} |C_{ij}|$). Обоснованием целесообразности такой постановки задачи является то, что (2.49) обеспечивает монотонность уменьшения по s структурной составляющей среднего значения критерия регулярности $(\overline{AR}(s) = E[AR(s)] = AR^b(s) + AR^v(s))$, где шумовая компонента $AR^v(s)$ монотонно возрастает по s при любом разбиении). Однако монотонность сохраняется и при некотором рассогласовании равенства (2.49) [11].

Очевидно, что процедура (2.63) в полном объеме трудно реализуема из-за необходимости полного перебора вариантов, в особенности при большом числе точек

$$L = \sum_{i=m}^{n-m} C_n^i$$

(число всех возможных вариантов), поэтому следует применять целесообразные приемы сокращения перебора. Из таких приемов сформулируем три:

- 1) задается некоторое небольшое число (например, 10) способов разбиения на основе других соображений, и к ним применяется процедура (2.63);
- 2) задается число точек разбиения n_A, n_B ($n = n_A + n_B$), тогда сравнивается только $C_n^{n^*}$ вариантов, где $n^* = \min(n_A, n_B)$;
- 3) выполняется квазиоптимальная двухэтапная процедура [20], когда сравниваются не все элементы матриц в (2.63), а только диагональные (с учетом того, что в матрице $X^T X$ недиагональные элементы всегда строго меньше диагональных), и отбирается некоторое число вариантов, из которых определяется лучший по полной процедуре (2.63). Рассмотрим несколько подробнее третий подход.

Если выборка $W=(Xy)$ получена в результате наблюдений за обычным режимом работы объекта, то можно говорить только о некотором приближении к ρ_B^2 - пропорциональному разбиению. Поскольку при этом задавать ρ_B^2 заранее не имеет смысла, можно применить предложенный в [20] подход для его вычисления, а тем самым и для приближенного выбора разбиения. Обозначим элемент информационной матрицы $X^T X$ через $\chi_{ij} = x_i^T x_j, i, j = \overline{1, m}$. Для некоторого разбиения матрицы X на подматрицы X_A, X_B введем величину соотношения соответствующих элементов информационных матриц: $\rho_{ij} = \chi_{ij}^A / \chi_{ij}^B$.

Ввиду симметричности рассматриваемых матриц всего имеется $m(m+1)/2$ различных значений коэффициентов ρ_{ij} . Предположив для определенности, что все χ_{ij}^A, χ_{ij}^B , положительны, введем следующую меру степени рассогласования информационных матриц:

$$\delta = (\rho_{\max} - \rho_{\min}) / (\rho_{\max} + \rho_{\min}),$$

где ρ_{\min}, ρ_{\max} - наименьшее и наибольшее из всех значений ρ_{ij} . Тогда очевидно, что в качестве наилучшего (в смысле ρ_B^2 -пропорциональности) разбиения из L заданных вариантов следует выбрать разбиение с номером

$$r^* = \arg \min_{r=1,L} \delta_r. \quad (2.64)$$

Если окажется, что полученное значение δ_{r^*} достаточно близко к нулю, то соответствующее разбиение принимается в качестве ρ_B^2 -пропорционального, и при этом автоматически определяется значение $\rho_B^2 \approx \rho_{\min} \approx \rho_{\max}$. Отметим, что принятая нормировка необходима ввиду разномасштабности величин ρ_{ij} для различных разбиений. Объем вычислений при полном решении задачи (2.64) может оказаться значительным при достаточно больших n и r_{\max} , однако его можно существенно уменьшить, учитывая только $\rho_{ii}, i = \overline{1, m}$, т.е. только диагональные члены матриц $X_A^T X_A$ и $X_B^T X_B$ для предварительного отбора нескольких лучших разбиений. Это следует из полученного в [20] неравенства, связывающего ρ_{ij} с ρ_{ii} и ρ_{jj} :

$$\rho_{ij} k_{ij}^A < \frac{1}{2} (\rho_{ii} + \rho_{jj}). \quad (2.65)$$

где введен косинус угла между векторами $x_i, x_j : k_{ij}^A = \chi_{ij}^A / \sqrt{\chi_{ii}^A \chi_{jj}^A}$.

Неравенство (2.65) позволяет контролировать значения $\rho_{ij}, i \neq j$, через значения $\rho_{ii}, i = \overline{1, m}$. Поэтому для сокращения объема вычислений можно предложить двухэтапную процедуру поиска лучшего разбиения: сначала (2.64) применяется только к диагональным элементам матриц $(X_A^T X_A)_r$ и $(X_B^T X_B)_r$ для приближенного отбора небольшого числа подходящих разбиений, а затем эти варианты анализируются на согласованность всех элементов указанных матриц с выбором лучшего разбиения.

Рассмотренный подход к решению проблемы разбиения позволяет максимально приблизиться к ρ_B^2 -пропорциональному (квадратично зависящему) разбиению, при котором условия применения критерия регулярности (и других критериев точности) являются наилучшими с точки зрения выбора прогнозирующей модели.

Связь квазиоптимального разбиения с традиционным разбиением "по дисперсии". Важно то, что предложенная выше квазиоптимальная процедура разбиения выполняется до начала алгоритма структурной идентификации. Поскольку это же свойство характеризует и традиционно применяемый в алгоритмах МГУА способ разбиения "по дисперсии точек наблюдений" [22], представляет интерес установить связь между этими двумя разбиениями.

Для матрицы плана $X = \{x_{ji}\}$, $i = \overline{1, m}$, $j = \overline{1, n}$, процедуру разбиения по дисперсии можно представить так: для каждого регрессора находится среднее

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ji}, \quad i = \overline{1, m},$$

от которого вычисляется величина разброса ("дисперсии") для каждой точки:

$$\eta_j^2 = \sum_{i=1}^m (x_{ji} - \bar{x}_i)^2, \quad j = \overline{1, n},$$

и все полученные значения η_j^2 ранжируются в порядке убывания. Затем "ранжированные точки" относятся в подвыборки A и B двумя различными способами: либо через одну точку (получение "подобных" частей), либо подряд (получение "непохожих" частей).

Этот способ разбиения в терминах квадратичной зависимости подвыборок можно интерпретировать так. Пусть выборка такова, что все $\bar{x}_i = 0$, тогда

$$\tilde{\eta}_j^2 \triangleq \eta_j^2|_{\bar{x}=0} = \sum_{i=1}^m (x_{ji})^2.$$

Далее, легко получить, что

$$\text{tr} X^T X = \sum_{i=1}^m \chi_{ii} = \sum_{i=1}^m \sum_{j=1}^n (x_{ji})^2 = \sum_{j=1}^n \tilde{\eta}_j^2,$$

следовательно, разбиение "по дисперсии через одну точку" - это попытка получить подвыборки, связанные соотношением $\text{tr} X_A^T X_A \approx \text{tr} X_B^T X_B$, и при этом $n_A \approx n_B$, или $\rho_B^2 = n_B / n_A \approx 1$. В этом случае есть основание ожидать наилучших условий для критериев группы точности. При разбиении же ранжированных точек подряд достигается цель $|\text{tr} X_A^T X_A - \text{tr} X_B^T X_B| \rightarrow \max$, что, как отмечалось выше, целесообразно при использовании критериев согласованности.

Таким образом, процедура ранжировки точек выборки данных "по дисперсии" может рассматриваться как еще один - причем наиболее простой по сравнению с предложенными выше - способ приближения к ρ_B^2 -пропорциональному разбиению. При этом, однако, следует иметь в виду, что условие $\text{tr} X_A^T X_A = \text{tr} X_B^T X_B$ является необходимым, но не достаточным для выполнения условия квадратичной зависимости $X_A^T X_A = X_B^T X_B$. Поэтому данный способ следует рассматривать как один из вариантов генерирования приближенно пропорциональных разбиений, наряду с рассмотренными выше более сложными (в смысле вычислительных затрат) вариантами.

3. МОДЕЛИРОВАНИЕ И ПРОГНОЗИРОВАНИЕ МАКРОЭКОНОМИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ МГУА

3.1. Методика выбора наиболее существенных аргументов моделей

При большом числе исходных аргументов построение модели занимает очень много машинного времени. В то же время практика построения моделей показывает, что среди представленных аргументов много несущественных, которые можно без ущерба для результирующей модели исключить. Ниже рассматривается один из способов решения этой проблемы с применением предварительных расчетов по многорядному алгоритму МГУА. Следует при этом иметь в виду, что разделение аргументов на существенные и несущественные справедливо для заданного набора данных и при изменении выборки (появлении новых данных) может изменяться.

Напомним процедуру селекции моделей в многорядном алгоритме МГУА [30]. В основу алгоритма положен целенаправленный перебор частных описаний с постепенным их усложнением при переходе от ряда к ряду. Частное описание $Y_{ij}^S = f(Y_i^{S-1}, Y_j^{S-1})$ - линейная или квадратичная функция двух промежуточных переменных предыдущего ряда ($i, j = 1, \dots, NA$; $S = 1, 2, \dots$; $Y_i^0 = X_i$; NA - число исходных аргументов). Из предыдущего ряда в следующий пропускается NA промежуточных переменных, удовлетворяющих критерию селекции $K_i^S = \min_{j=1, NA} K_{ij}^S$ (K_{ij}^S - корреляционное отношение Y_{ij}^S и Y).

При таком способе отбора на каждом ряду сохраняется исходный базис, что позволяет восстанавливать искомую модель. Следовательно, и на последнем ряду также будет NA моделей, т.е. свобода выбора в данном случае равна NA . Из сказанного следует, что в каждой i -й промежуточной переменной обязательно будет присутствовать исходный аргумент X_i , к которому оптимальным образом подбираются наилучший для данного случая ансамбль из остальных исходных аргументов. Следовательно, можно построить структурную таблицу размером $NA * NA$, i -я строка которой соответствует i -й модели, а j -й столбец - j -тому аргументу ($i, j = 1, \dots, NA$), ij -я ячейка этой таблицы равна j , если j -й аргумент участвует в формировании i -й модели, или 0, если не участвует.

Такая таблица показывает, какие исходные аргументы участвуют в формировании каждой i -й модели. Далее подсчитываем, сколько раз участвует в формировании моделей каждый исходный аргумент (частота появления j -го аргумента во всех моделях), т.е. определяем значения $NumX_j$. Эти числа лежат в интервале от 1 до NA . Если $NumX_j = 1$, то можно утверждать, что j -й аргумент несуществен в данном случае, так как присутствует только там, где он насильно удерживается в силу изложенного выше алгоритма отбора лучших моделей. Если $NumX_j = NA$, то можно утверждать, что j -й аргумент существен для данного случая. Это крайние случаи, поэтому следует ввести некоторый порог для определения существенности j -го аргумента: например, будем считать аргумент существенным, если $NumX_j > 0.5 * NA$. Кроме того можно, упорядочить аргументы по убыванию величины $NumX_j$, и пользуясь дополнительными критериями, выбрать столько аргументов, сколько может быть допустимо в данном конкретном случае.

Например, исходная таблица данных для моделирования объема производства легкой промышленности, взятая из [31] (см.ниже), содержит 23 исходных аргумента ($NA = 23$):

- X_1 - доходы населения (млрд.грн.),
- X_2 - индекс потребительских цен (%),
- X_3 - индексы ВВП (%),
- X_4 - объем производства промышленности, всего (млрд.грн.),
- X_5 - розничный товароборот всего (млрд.грн.),
- X_6 - индексы зарплат (реальная, %),
- X_7 - средние банковские процентные ставки на кредиты (%),
- X_8 - официальный курс гривни к доллару США (грн.),
- X_9 - налог на добавленную стоимость (млрд.грн.),
- X_{10} - налог на прибыль предприятий (млрд.грн.),
- X_{11} - расходы консолидированного бюджета, всего (млрд.грн.),
- X_{12} - расходы бюджета на народное хозяйство (млрд.грн.),
- X_{13} - индекс оптовых цен легкой промышленности (%),
- X_{14} - занятость в легкой промышленности (млн.человек),
- X_{15} - общая занятость (млрд.человек),
- X_{16} - заработная плата (номинальная, грн.),
- X_{17} - учетная ставка для коммунальных банков(%),
- X_{18} - денежная масса наличности в обращении (млрд.грн.),
- X_{19} - дебиторская задолженность (млрд.грн.),
- X_{20} - кредиторская задолженность (млрд.грн.),
- X_{21} - акцизный сбор (млн.грн.),
- X_{22} - налог на доход граждан (млн.грн.),
- X_{23} - доход консолидированного бюджета (млрд.грн.),
- Y - Объем производства легкой промышленности (млрд.грн.)

В результате выполнения процедуры построения моделей по многорядному алгоритму была получена приведенная ниже структурная таблица.

1	2	0	4	0	6	0	8	0	10	11	0	13	0	0	0	0	0	20	0	0	0	
1	2	3	4	0	6	0	0	0	10	0	12	0	14	0	16	17	0	0	0	0	0	
1	2	3	4	0	6	7	0	0	10	0	12	0	14	0	16	17	0	0	20	21	0	0
1	2	0	4	0	6	0	0	0	0	11	12	13	14	0	0	17	18	19	0	0	0	0
1	2	3	4	5	6	7	0	0	0	11	12	13	14	0	16	17	0	19	0	0	0	0
1	0	0	4	0	6	0	8	0	10	11	12	13	14	0	16	0	0	0	20	0	0	0
1	0	3	4	0	6	7	8	0	10	11	12	13	0	0	0	17	0	19	0	21	22	0
1	0	0	4	5	6	7	8	0	10	11	12	13	14	0	16	17	18	19	0	0	0	0
1	0	3	4	0	6	7	8	9	10	11	12	13	14	0	16	17	0	19	0	0	0	0
1	2	0	4	0	6	7	0	0	10	0	0	0	14	0	16	17	0	19	20	0	0	0
1	2	3	4	0	6	7	0	0	10	11	12	0	0	0	16	0	0	19	20	0	0	0
1	0	3	4	0	6	0	8	0	10	11	12	13	14	0	16	17	0	19	0	0	22	0
1	2	0	4	5	6	7	0	0	0	11	0	13	14	0	0	17	18	19	0	0	0	0
1	0	3	4	0	6	7	0	0	10	0	12	13	14	0	16	17	0	0	0	0	0	0

1	2	3	4	0	6	7	8	0	10	11	0	13	14	15	16	17	18	19	0	0	22	0
1	2	0	4	0	6	0	8	0	10	11	0	13	14	0	16	17	0	19	0	0	0	0
1	0	0	4	0	6	7	8	0	10	11	12	13	14	15	16	17	0	19	20	0	0	0
0	2	0	4	0	6	7	0	0	0	11	0	0	14	0	0	17	18	0	20	0	0	0
1	0	0	4	0	6	7	8	0	10	11	12	13	14	15	16	17	0	19	0	21	0	0
1	0	0	4	0	6	7	8	0	10	11	12	13	14	15	16	17	0	19	20	21	22	0
1	2	3	4	0	6	7	8	0	10	11	12	13	0	0	16	17	0	19	0	21	0	0
1	2	0	4	0	6	7	0	0	0	11	12	0	14	0	16	0	18	0	0	0	22	0
1	2	0	4	0	6	0	0	0	10	0	0	0	0	0	16	17	0	19	20	0	0	23
22	14	10	23	3	23	16	12	1	18	18	16	16	18	4	18	19	6	16	9	5	5	1

Последняя строка таблицы показывает частоту участия каждого аргумента в формировании всех моделей. Если воспользоваться порогом $NumX_j > 0.5 * NA$, получим, что для моделирования достаточен набор из следующих 13-ти аргументов: $X_1, X_2, X_4, X_6, X_7, X_{10}, X_{11}, X_{12}, X_{13}, X_{15}, X_{16}, X_{17}, X_{19}$.

Пример 1: поступления в государственный бюджет Украины

Для этой задачи были выбраны следующие аргументы (параметры) [31]: X_1 – денежные доходы населения (млрд.грн.); X_2 – индекс потребительских цен (%); X_3 – индекс оптовых цен в промышленности (%); X_4 – индекс ВВП (%); X_5 – объем производства промышленности (млрд.грн.); X_6 – розничный товароборот (млрд.грн.); X_7 – общая занятость населения (млн.чел.); X_8 – зарплата (номинальная, грн.); X_9 – индекс зарплат (реальная, %); X_{10} – учетная ставка коммерческих банков (%); X_{11} – процентные ставки на кредиты (%); X_{12} – денежная масса наличности в обращении (млрд.грн.); X_{13} – официальный курс гривны к доллару США (грн. за 100 дол.); X_{14} – дебиторская задолженность (млрд.грн.); X_{15} – кредиторская задолженность (млрд.грн.); X_{16} – налог на добавленную стоимость (млн.грн.); X_{17} – налог на прибыль юридических лиц (млн.грн.); X_{18} – акцизный сбор (млн. грн.); X_{19} – налог на прибыль с граждан (млн. грн.); X_{20} – расходы консолидированного бюджета, всего (млрд. грн.); X_{21} – расходы консолидированного бюджета на народное хозяйство (млн. грн.); Y_1 – доход консолидированного бюджета (млрд. грн.).

Значения перечисленных переменных за период с ноября 1995 по июнь 1997 гг. представлены в таблице 3.1 (длина исходной выборки – 20 точек). По этим данным получена модель (M1) зависимости дохода бюджета от текущих значений аргументов, причем две последние точки составили экзаменационную последовательность, т.е. для моделирования применялись только восемнадцать точек.

Построенная модель имеет вид:

$$V_{1,t} = 0.4063 - 1.0393 \times X_{6,t} - 0.0003 \times X_{16,t} + 0.0013 \times X_{17,t} + 0.0063 \times X_{19,t} + 0.5555 \times X_{20,t}, \quad (3.1)$$

где t – номер точки, V_1 – модельная оценка выходной переменной Y_1 .

Модель имеет следующие характеристики:

$$СКО = 0.1528; \quad R = \max_{t=1,20} (V_{1,t} - Y_{1,t}) = 0.2959; \quad S = R / (Y_{1,max} - Y_{1,min}) * 100\% = 11.62\%,$$

где СКО – величина среднеквадратичного отклонения оценок V_t от табличных значений $Y_{l,t}$, R – наибольшая абсолютная ошибка, S – относительная максимальная ошибка в процентах от “размаха” значений $Y_{l,t}$.

В таблице 3.2 приведены оценки дохода бюджета (М1), полученные по модели (3.1), на рисунке 3.1 представлен график их изменения. Как видим, модель имеет неплохую точность, однако попытаемся еще ее повысить. Для этого расширим исходный базис, введя дополнительные переменные – функции от исходных аргументов, в данном случае запаздывающие (лаговые) значения от выходной величины и исходных аргументов.

В результате экспериментов по описанной выше процедуре были выбраны следующие дополнительные аргументы: $X_{22,t}=Y_{1,t-4}$; $X_{23,t}=X_{6,t-1}$; $X_{24,t}=X_{16,t-1}$; $X_{25,t}=X_{17,t-1}$; $X_{26,t}=X_{17,t-2}$. При этом длина выборки уменьшилась на 4 точки и составляет 16 точек, из которых две предпоследние – экзаменационные. С учетом этого была получена модель (М2):

$$V_{1,t} = 1.6285 - 0.0685 \times X_{5,t} - 0.0535 \times X_{6,t} - 0.0152 \times X_{9,t} + \\ + 0.0012 \times X_{17,t} + 0.7936 \times X_{20,t} + 0.0001 \times X_{17,t-1}. \quad (3.2)$$

Модель имеет следующие характеристики:

СКО = 0.1223; $R = 0.1259$; $S = 8.87\%$.

В таблице 3.2 приведены оценки дохода бюджета (М2 по модели (3.2), на рисунке 3.1 - график их изменения. Эта модель имеет хорошие показатели точности на всей выборке, т.е. введение дополнительных аргументов существенно повысило адекватность модели.

Таблица 3.А – Исходные данные для моделирования дохода бюджета

п/н	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	2.981	106.2	104.2	88.0	5.7495	1.1613	15.082	103.29	106.4	89.0	116.4
2	3.504	104.6	102.6	88.2	5.8579	1.2877	14.906	134.38	110.5	103.8	105.8
3	2.459	109.4	103.4	87.4	5.3663	1.1218	14.743	114.52	79.2	105.0	110.5
4	2.931	107.4	102.9	88.5	5.5110	1.1840	14.718	120.93	98.2	105.0	110.9
5	2.888	103.0	102.9	91.4	6.0512	1.2761	14.594	129.33	102.3	97.1	102.5
6	3.185	102.4	101.5	90.4	6.0128	1.2628	14.509	129.59	98.6	76.3	91.2
7	3.083	100.7	100.8	90.9	5.6732	1.2286	14.409	132.10	100.3	67.7	85.7
8	2.943	100.1	100.4	91.4	5.3162	1.2103	14.352	137.74	103.7	52.6	73.4
9	3.708	100.1	100.6	90.8	5.3933	1.3172	14.191	143.73	104.4	40.0	66.2
10	3.265	105.7	100.4	90.5	5.2719	1.3909	14.074	143.05	94.1	40.0	63.6
11	3.721	102.0	101.0	89.9	5.2501	1.3128	13.997	146.09	100.0	40.0	63.2
12	3.622	101.5	100.1	90.0	5.7942	1.3213	13.879	148.21	99.6	40.0	63.7
13	3.167	101.2	101.2	89.9	5.6521	1.2533	13.768	144.76	96.6	40.0	66.3
14	4.281	100.9	100.9	90.0	5.8947	1.4177	13.646	163.69	111.8	40.0	61.4
15	3.283	102.2	100.4	89.6	5.1665	1.1640	13.271	139.50	84.6	37.7	63.7
16	3.462	101.2	100.4	91.1	5.0760	1.1939	13.361	138.90	98.1	35.0	62.6
17	3.938	100.1	100.6	91.5	5.6491	1.4516	13.319	147.73	105.7	25.0	53.2
18	4.144	100.8	100.9	92.3	5.5330	1.4679	13.214	146.75	98.2	25.0	52.4
19	3.885	100.8	100.3	92.6	5.3377	1.4633	13.102	153.39	103.7	21.0	53.7
20	4.140	100.1	100.4	92.5	5.3205	1.5080	13.064	158.33	102.9	21.0	50.3

Продолжение таблицы 3.1

н/п	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	Y1
1	2	3	4	5	6	7	8	9	10	11	12
1	2.226	177.88	22.0021	28.2824	503.4	485.4	33.0	211.1	2.4849	331.6	2.3069
2	2.623	179.50	22.2502	30.5431	543.7	662.4	30.8	185.1	3.5769	535.6	2.7199
3	2.575	182.80	25.4619	33.8009	431.3	208.9	26.5	172.9	2.2005	87.6	1.6739
1	2	3	4	5	6	7	8	9	10	11	12
4	2.799	188.10	28.8513	37.2235	393.4	467.3	32.3	186.4	2.2289	119.4	2.0984
5	2.800	189.15	35.5120	45.1525	472.0	608.0	34.6	184.1	3.2165	241.2	2.7323
6	3.077	187.27	33.6751	45.8835	508.7	375.8	34.4	206.0	2.2839	214.2	2.1540
7	3.190	184.19	34.0136	45.9351	428.2	371.4	32.8	210.9	2.2529	184.3	2.0774
8	3.324	181.98	37.3405	55.5831	413.3	390.0	30.7	193.6	2.4845	207.7	2.0104
9	3.617	178.08	37.1381	53.8383	611.3	490.7	91.7	247.7	3.1284	233.2	2.7329
10	3.190	176.00	38.5272	55.7474	508.2	393.6	55.8	231.7	2.2583	139.0	2.3433
11	3.330	176.00	40.6251	60.2888	424.7	438.0	69.0	216.9	2.6374	242.1	2.3604
12	3.471	177.48	42.9449	59.2260	648.0	425.1	88.4	271.0	3.3818	430.8	2.8846
13	3.499	185.92	43.1946	60.5737	559.8	525.1	74.2	214.9	3.0958	461.8	2.8536
14	4.041	188.39	48.0181	73.1681	894.2	756.9	81.8	303.2	4.5902	888.1	4.2207
15	4.087	189.14	53.2016	72.1561	460.5	147.8	57.6	218.9	2.1049	178.4	1.8246
16	4.143	184.36	49.9248	72.7840	406.3	296.8	56.8	220.1	2.3274	72.3	2.0475
17	4.306	183.71	59.3116	84.9681	631.1	412.1	72.0	247.2	3.1551	277.7	2.6655
18	4.763	184.79	60.1183	86.8049	617.7	382.9	77.2	268.8	2.4494	576.5	2.0399
19	4.847	184.28	60.1171	86.1344	651.7	420.7	81.9	243.9	2.3784	483.1	2.0122
20	5.102	185.78	65.2506	95.1487	659.6	387.9	85.2	265.8	2.5684	229.6	2.1322

Таблица 3.В — Оценки доходов бюджета

п/п	Y1	V1	
		M1	M2
1	2.3069	2.2961	—
2	2.7199	3.0158	—
3	1.6739	1.8088	—
4	2.0984	2.0209	—
5	2.7323	2.7265	2.9201
6	2.1540	2.0584	2.0161
7	2.0774	2.0176	1.9584
8	2.0104	2.1310	2.1451
9	2.7329	2.7717	2.7620
10	2.3433	2.0790	2.1317
п/п	Y1	V1	
		M1	M2
11	2.3604	2.2880	2.3823
12	2.8846	2.9367	2.9358
13	2.8536	2.7019	2.8353
14	4.2207	4.0713	4.0909
15	1.8246	1.8043	1.8931
16	2.0475	2.0217	1.9860
17	2.6655	2.7018	2.6841
18	2.0399	2.2944	2.2071
19	2.0122	2.2463	2.1310
20	2.1322	2.3725	2.2753

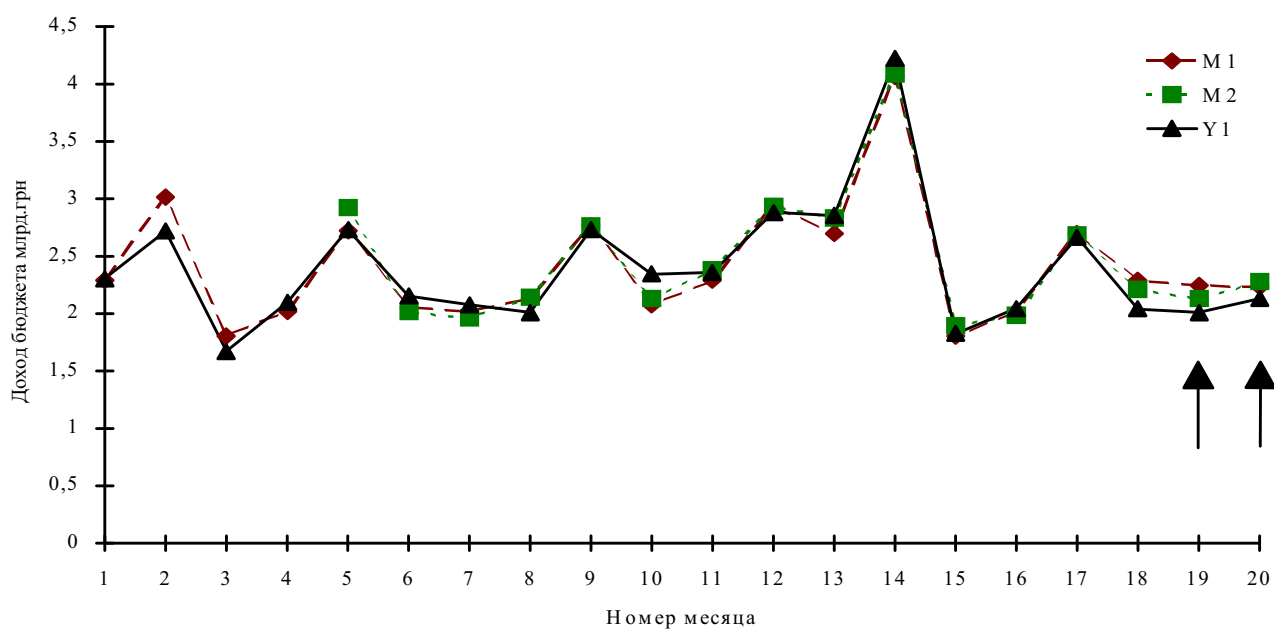


Рисунок 3.1 — График изменения доходов бюджета (Y_I) и его оценок (M1 и M2) по формулам (3.1 и 3.2) (стрелками отмечены экзаменационные точки).

Пример 2: объем производства легкой промышленности

Как показано выше в подразделе 3.1, в этой задаче из 23 начальных переменных для моделирования были отобраны следующие 13 наиболее существенных для данной выборки аргумента (для удобства изменяем нумерацию): X_1 – денежные доходы населения (млрд.грн.); X_2 – индекс потребительских цен (%); X_3 – объем производства промышленности, всего (млрд.грн.); X_4 – индексы зарплаты (реальная, %); X_5 – налог на добавленную стоимость (млрд.грн.); X_6 – налог на прибыль предприятий (млрд. грн.); X_7 – расходы консолидированного бюджета, всего (млрд.грн.); X_8 – расходы бюджета на народное хозяйство (млрд.грн.); X_9 – занятость в легкой промышленности (млн.чел.); X_{10} – общая занятость (млрд.чел.); X_{11} – зарплата (номинальная, грн.); X_{12} – денежная масса наличности в обращении (млрд.грн.); X_{13} – кредиторская задолженность (млрд.грн.).

Входной величиной Y_2 является объем производства легкой промышленности в млрд.грн. В таблице 3.3 приведены значения перечисленных аргументов и выходной величины за период времени с ноября 1995 по июнь 1997 гг. [31], т.е. длина исходной выборки – 20 точек. Отметим, что показатели X_1 , X_2 и X_4 характеризуют уровень жизни населения, X_5 – X_8 являются бюджетными, X_9 – отраслевой, X_{12} – финансовый, а остальные являются макроэкономическими.

Вначале рассмотрим построение *экстраполирующей* модели объема производства легкой промышленности. Под ней понимается модель, в которой текущее значение выходной величины зависит от текущих значений аргументов и, возможно, от прошлых значений аргументов и выходной величины.

По данным таблицы 3.3 получена модель зависимости объема производства легкой промышленности от текущих значений аргументов (М3), причем две последние точки выборки составляли экзаменационную последовательность, т.е. для построения модели использовались только первые восемнадцать точек. Построенная модель имеет вид:

$$V_{2,t} = 0.3789 + 0.0467 \times X_{3,t} + 0.001 \times X_{4,t} - 0.0023 \times X_{7,t} - 0.0024 \times X_{10,t} - 0.0007 \times X_{11,t} - 0.0017 \times X_{13,t}, \quad (3.3)$$

где t – номер точки, $V_{2,t}$ – оценка величины $Y_{2,t}$ по модели.

Модель имеет следующие характеристики:

СКО = 0.00715; $R = 0.01679$; $S = 18.06\%$,

В таблице 3.3 приведены оценки выходной величины (М3), полученные по модели (3.3), на рисунке 3.2 представлен график изменения оценок выходной величины. Как видно из этих результатов, модель, построенная только на основании исходного базиса переменных, имеет достаточно хорошую точность даже на экзаменационных точках, т.е. в режиме экстраполяции, поэтому для нее не предпринимались попытки увеличения точности при помощи расширения базиса введением дополнительных переменных – функций от исходных аргументов.

Рассмотрим теперь задачу построения *прогнозирующей* модели, т.е. модели, в которой будущие значения выходной величины зависят от текущих значений аргументов и

выходной величины и, возможно, от их прошлых значений. В данном случае выходной величиной является будущее значение объема производства легкой промышленности $Y_{2,t+1}$, а число аргументов увеличивается на единицу $X_{14,t} \equiv Y_{2,t}$.

По данным таблицы 3.3 построена прогнозирующая на один шаг вперед модель (М4) зависимости будущих значений объема производства легкой промышленности от текущих значений аргументов и выходной величины (исходный базис). Для построения модели использовались только первые семнадцать точек; точки 18 и 19 составляли экзаменационную выборку, а 20-я точка используется для прогноза 21-го значения выходной величины. Модель имеет вид:

$$V_{2,t+1} = 0.1064 - 0.0210 \times X_{3,t} - 0.0307 \times X_{6,t} + 0.0062 \times X_{9,t} + 0.0047 \times X_{10,t} + 0.5058 \times Y_{2,t}, \quad (3.4)$$

Она имеет следующие характеристики:

$$\text{СКО} = 0.01027; \quad R = 0.02127; \quad S = 27.44\%.$$

В таблице 3.4 приведены оценки выходной величины (М4), полученные по модели (3.4), на рисунке 3.2 представлен график изменения оценок выходной величины. Как видно из приведенных результатов, модель имеет невысокую точность. Для ее повышения расширим исходный базис, введя запаздывающие значения аргументов. Путем экспериментов были установлены следующие дополнительные переменные: $X_{15,t} = X_{8,t-1}$; $X_{16,t} = X_{8,t-2}$; $X_{17,t} = X_{8,t-1}$; $X_{18,t} = X_{9,t-1}$; $X_{19,t} = X_{9,t-2}$; $X_{20,t} = X_{9,t-3}$; $X_{21,t} = Y_{2,t}$; $X_{22,t} = Y_{2,t-1}$; $X_{23,t} = Y_{2,t-2}$; $X_{24,t} = X_{7,t-1}$; $X_{25,t} = X_{7,t-2}$; $X_{26,t} = X_{7,t-3}$, при этом длина выборки уменьшилась на 3 точки и составляет 17 точек, из которых последняя используется для вычисления будущей оценки выходной величины, а две предпоследние являются экзаменационными. С учетом этого была получена модель (М4) вида:

$$V_{2,t+1} = 0.0975 + 0.0018 \times X_{2,t} - 0.0038 \times X_{3,t} - 0.0007 \times X_{4,t} + 0.0049 \times X_{9,t} + 0.0039 \times X_{10,t} + 0.0003 \times X_{12,t} - 0.0003 \times X_{13,t} + 0.2930 \times Y_{2,t} - 0.0062 \times X_{9,t-1} + 0.0011 \times X_{9,t-2} - 0.3352 \times Y_{2,t-1} - 0.0584 \times Y_{2,t-2} - 0.0075 \times X_{7,t-1}, \quad (3.5)$$

Модель имеет следующие характеристики:

$$\text{СКО} = 0.00202; \quad R = 0.00638; \quad S = 6.41\%.$$

В таблице 3.4 приведены оценки выходной величины (М5), полученные по модели (3.5), на рисунке 3.2 представлен график изменения оценок выходной величины.

Из приведенных результатов следует, что мы получили прогнозирующую модель, имеющую достаточно хорошую точность на всех точках, используемых как для построения модели, так и для экзамена, поэтому можно предположить, что и прогнозное значение тоже достаточно хорошо совпадет с реальным.

Для проверки адекватности выбранного выше состава дополнительных аргументов по данным 1995-96 гг. (10 точек, с учетом запаздываний) была построена модель для вычисления прогноза на первое полугодие 1997 года (М6):

$$V_{2,t+1} = 0.3257 - 0.0009 \times X_{2,t} - 0.0047 \times X_{3,t} + 0.0025 \times X_{9,t} + 0.0008 \times X_{12,t} + 0.6616 \times Y_{2,t} + 0.0069 \times X_{8,t-1} - 0.6547 \times Y_{2,t-1}, \quad (3.6)$$

Модель имеет следующие характеристики:

$$\text{СКО} = 0.00707; \quad R = 0.01573; \quad S = 21.73\%.$$

Как и в предыдущем случае, последняя точка используется для вычисления оценки объема производства легкой промышленности в июне 1997 г., а предпоследние 6 точек были отнесены в экзаменационную выборку.

В таблице 3.4 приведены оценки выходной величины (M_6), полученные по модели (3.6), на рисунке 3.3 представлен график изменения оценок выходной величины.

Из приведенных результатов следует, что даже в этом случае мы получили прогнозирующую модель, имеющую достаточно хорошую точность на всех точках, используемых как для построения модели, так и для экзамена, причем, как отмечалось выше, в экзаменационную последовательность были отнесены 6 предпоследних точек выборки данных. Это говорит о надежности прогнозирования объема легкой промышленности на месяц вперед с применением МГУА.

Таблица 3.3 – Исходные данные для моделирования объема производства легкой промышленности

н/п	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	2.981	6.2	5.7495	6.4	16.4	0.4854	2.4849
2	3.504	4.6	5.8579	10.5	5.8	0.6624	3.5769
3	2.459	9.4	5.3663	-20.80	10.5	0.2089	2.2005
4	2.931	7.4	5.5110	-1.80	10.9	0.4673	2.2289
5	2.888	3.0	6.0512	2.3	2.5	0.6080	3.2165
6	3.185	2.4	6.0128	-1.40	-8.80	0.3758	2.2839
7	3.083	0.7	5.6732	0.3	-14.30	0.3714	2.2529
8	2.943	0.1	5.3162	3.7	-26.60	0.3900	2.4845
9	3.708	0.1	5.3933	4.4	-33.80	0.4907	3.1284
10	3.265	5.7	5.2719	-5.90	-36.40	0.3936	2.2583
11	3.721	2.0	5.2501	0.0	-36.80	0.4380	2.6374
12	3.622	1.5	5.7942	-0.40	-36.30	0.4251	3.3818
13	3.167	1.2	5.6521	-3.40	-33.70	0.5251	3.0958
14	4.281	0.9	5.8947	11.8	-38.60	0.7569	4.5902
15	3.283	2.2	5.1665	-15.40	-36.30	0.1478	2.1049
16	3.462	1.2	5.0760	-1.90	-37.40	0.2968	2.3274
17	3.938	0.1	5.6491	5.7	-46.80	0.4121	3.1551
18	4.144	0.8	5.5330	-1.80	-47.65	0.3829	2.4494
19	3.885	0.8	5.3377	3.7	-46.30	0.4207	2.3784
20	4.140	0.1	5.3205	2.9	-49.73	0.3879	2.5684

Продолжение таблицы 3.3

н/п	X8	X9	X10	X11	X12	X13	Y2
1	0.3316	4.0	15.082	103.29	89.0	22.0021	0.1623
2	0.5356	3.9	14.906	134.38	103.8	22.2502	0.1468
3	0.0876	2.5	14.743	114.52	105.0	25.4619	0.1106
4	0.1194	1.6	14.718	120.93	105.0	28.8513	0.1279
5	0.2412	3.5	14.594	129.33	97.1	35.5120	0.1378
6	0.2142	2.1	14.509	129.59	76.3	33.6751	0.1417
7	0.1843	0.5	14.409	132.10	67.7	34.0136	0.1096
8	0.2077	0.6	14.352	137.74	52.6	37.3405	0.0995
9	0.2332	1.6	14.191	143.73	40.0	37.1381	0.0938
10	0.1390	0.0	14.074	143.05	40.0	38.5272	0.0979
11	0.2421	1.0	13.997	146.09	40.0	40.6251	0.1022
12	0.4308	0.7	13.879	148.21	40.0	42.9449	0.1200
13	0.4618	2.3	13.768	144.76	40.0	43.1946	0.1065
14	0.8881	-0.01	13.646	163.69	40.0	48.0181	0.1093
15	0.1784	0.8	13.271	139.50	37.7	53.2016	0.0693
16	0.0723	0.8	13.361	138.90	35.0	49.9248	0.0917
17	0.2777	0.4	13.319	147.73	25.0	59.3116	0.0964
18	0.5765	-0.9	13.214	146.75	25.0	60.1183	0.0941
19	0.4831	-1.2	13.102	153.39	21.0	60.1171	0.0882
20	0.2296	-0.7	13.064	158.33	21.0	65.2506	0.0866

Таблица 3.С — Значения оценок выходной величины

н/п	Y2	V2			
		M3	M4	M5	M6
1	0.1623	0.1635	-	-	-
2	0.1468	0.1430	0.1483	-	-
3	0.1106	0.1130	0.1313	-	-
4	0.1279	0.1293	0.1278	-	-
5	0.1378	0.1366	0.1199	0.1336	0.1338
6	0.1417	0.1350	0.1204	0.1446	0.1484
7	0.1096	0.1180	0.1212	0.1099	0.1140
8	0.0995	0.1015	0.1019	0.0997	0.1009
9	0.0938	0.1031	0.1040	0.0954	0.0922
10	0.0979	0.1018	0.1019	0.0994	0.1028
11	0.1022	0.0942	0.0990	0.1043	0.1095
12	0.1200	0.1111	0.1061	0.1187	0.1142
13	0.1065	0.1049	0.1017	0.1068	0.1029
14	0.1093	0.1125	0.1042	0.1081	0.1044
15	0.0693	0.0723	0.0785	0.0685	0.0845
16	0.0917	0.0832	0.0955	0.0907	0.0989
17	0.0964	0.1020	0.1046	0.0972	0.1023
18	0.0941	0.0928	0.0887	0.0929	0.1098
19	0.0882	0.0868	0.0823	0.0836	0.0936
20	0.0866	0.0766	0.0799	0.0853	0.0871
21	-	-	0.0834	0.0866	0.0903

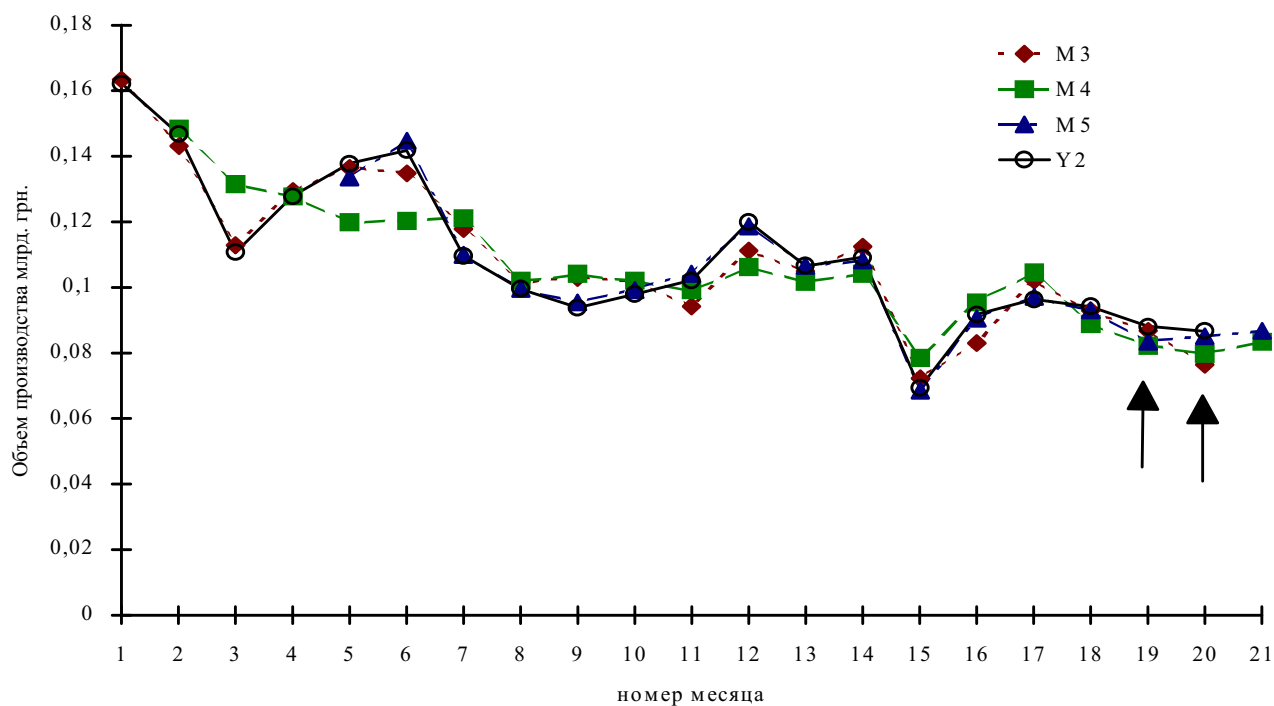


Рисунок 3.2 – Графики изменения объема производства (Y_2) и его оценок (M₃ - M₅) по моделям (3.3) - (3.5) (стрелками отмечены экзаменационные точки)

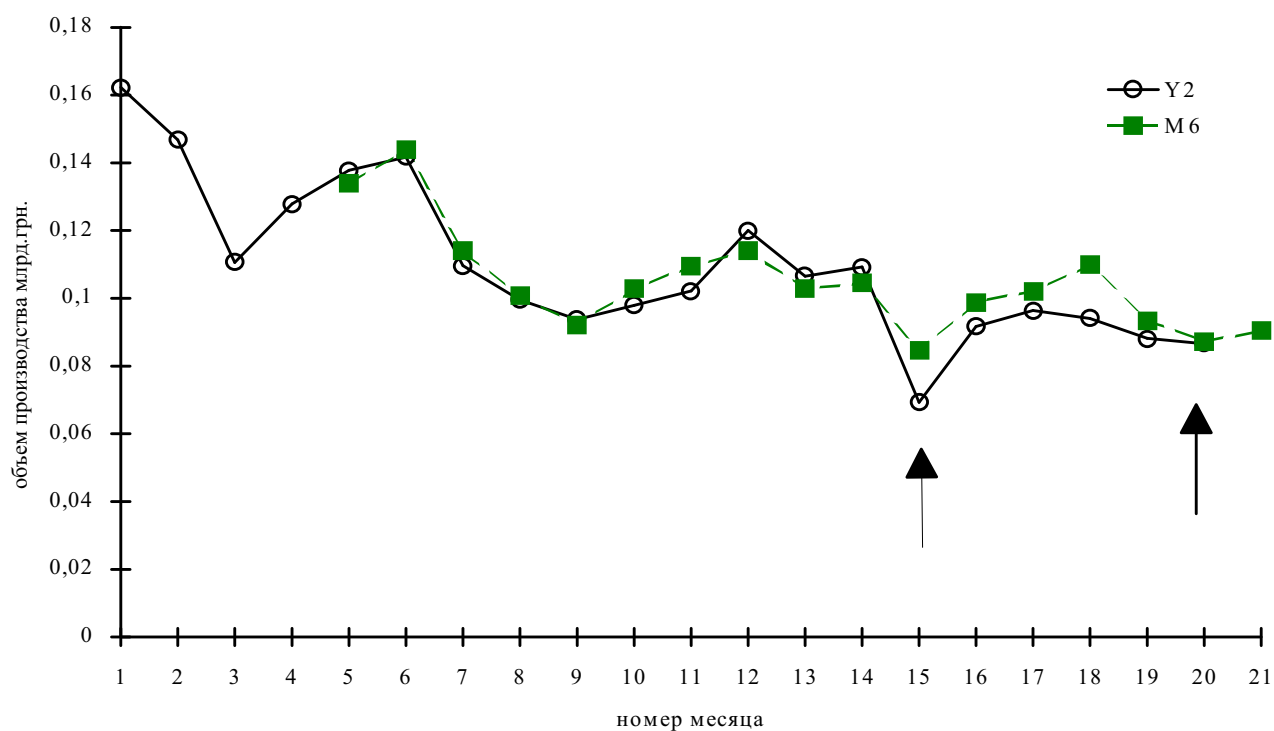


Рисунок 3.3 – График изменения объема производства (Y_2) и его оценок M₆ по модели (3.6) (стрелками отмечены начало и конец экзаменационной подвыборки)

Пример 3: процесс изменения индекса инфляции

Для построения модели по данным, взятым из [32], при помощи анализа структур моделей (см. 3.1), построенных по многорядному алгоритму МГУА, были отобраны наиболее существенные аргументы. Ими оказались: X_1 – накопления личные (\$ млн.); X_2 – число безработных всего; X_3 – процентные ставки (по Муди); X_4 – потребление личное (\$ млн.); X_5 – доходы личные (\$ млн.); X_6 – валовой национальный продукт; Y_3 – инфляция (рассчитывалась по формуле, приведенной в [32] на стр.186). Соответствующие данные приведены в таблице 3.5.

По данным таблицы 3.5 была получена модель (М7) зависимости инфляции от текущих значений аргументов, причем три последние точки выборки составляли экзаменационную последовательность, т.е. для построения модели были использованы только первые пятнадцать точек. Модель имеет вид:

$$V_{3,t} = 0.0004 \times X_{1,t} + 0.0109 \times X_{3,t} - 0.001 \times X_{4,t} + 0.0002 \times X_{5,t} - 0.0001 \times X_{6,t} \quad (3.7)$$

и имеет следующие характеристики:

$$\text{СКО} = 0.0087; \quad R = 0.0194; \quad S = 35.86\%.$$

В таблице 3.6 приведены оценки инфляции (М7) по модели (3.7), на рисунке 3.4 – график их изменения. Видно, что модель плохая как на обучающих и проверочных точках, так и на экзаменационных. Для повышения точности расширим базис, введя дополнительные переменные – запаздывающие значения от исходных аргументов.

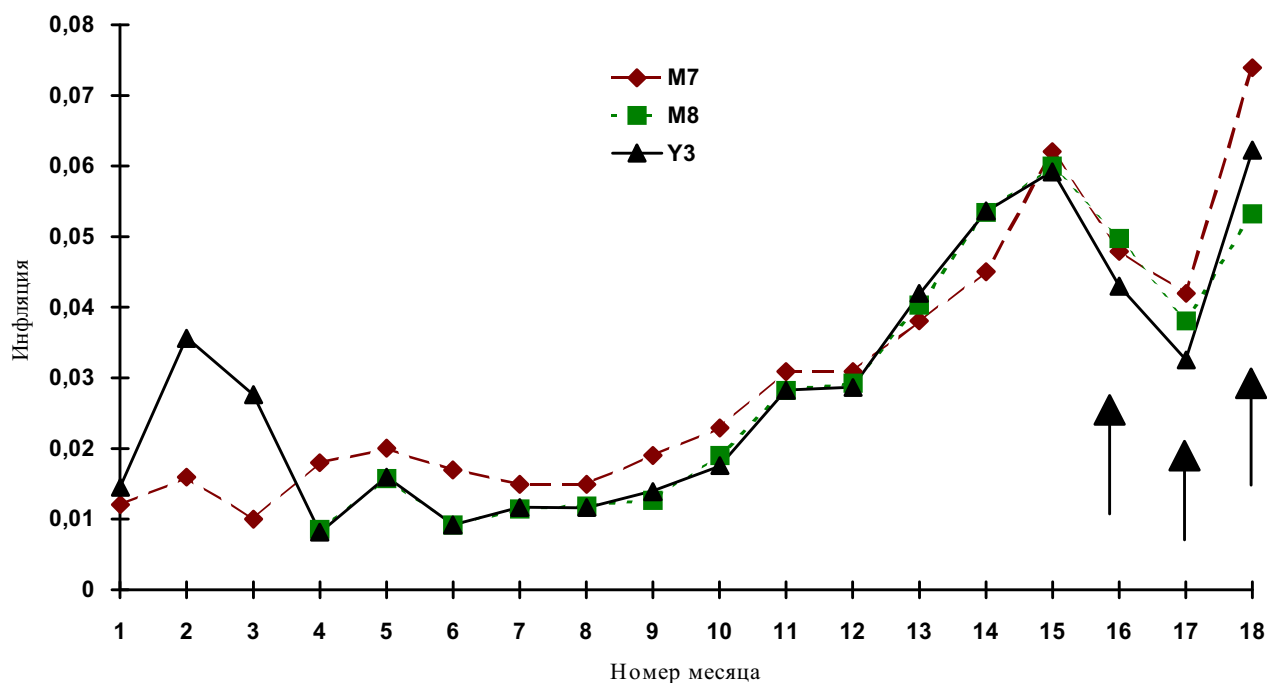
Таблица 3.D – Исходные данные для моделирования инфляции

п/п	X1	X2	X3	X4	X5	X6	Y3
1	2	3	4	5	6	7	8
1	36.0	4.1	2.77	38.9	333.0	446.1	0.0145
2	34.1	4.3	3.12	40.8	351.1	452.5	0.0357
3	33.0	6.8	2.15	37.9	361.2	447.3	0.0276
4	34.7	5.5	3.36	44.3	383.5	475.9	0.0081
5	32.3	5.5	3.53	45.3	401.0	487.7	0.0160
6	31.7	6.7	3.00	44.2	416.8	497.2	0.0092
7	37.1	5.5	3.00	49.5	442.6	529.8	0.0117
8	39.1	5.7	3.23	53.9	465.5	551.0	0.0116
9	49.5	5.2	3.55	59.2	497.5	581.0	0.0140
10	55.4	4.5	4.04	66.3	538.9	617.8	0.0175
1	2	3	4	5	6	7	8
11	65.1	3.8	4.50	70.8	587.2	658.1	0.0283
12	65.0	3.8	4.19	73.1	629.3	675.2	0.0287
13	68.3	3.6	5.17	84.0	688.9	706.6	0.0419
14	60.6	3.5	5.87	90.8	750.9	725.6	0.0536
15	76.2	4.9	5.95	91.3	808.3	722.5	0.0593
16	87.4	5.9	4.88	103.9	864.0	746.3	0.0430
17	97.9	5.6	4.50	118.4	944.9	792.5	0.0326
18	120.2	4.9	6.44	130.3	1055.0	839.2	0.0622

Таблица 3.Е – Оценки процесса инфляции

п/н	Y3	V3	
		M7	M8
1	0.0145	0.0125	–
2	0.0357	0.0163	–
3	0.0267	0.0107	–
4	0.0081	0.0187	0.0086
5	0.0160	0.0205	0.0157
6	0.0092	0.0174	0.0093
7	0.0117	0.0147	0.0114
8	0.0116	0.0152	0.0119
9	0.0140	0.0192	0.0127

п/н	Y3	V3	
		M7	M8
10	0.0175	0.0227	0.0189
11	0.0283	0.0307	0.0283
12	0.0287	0.0305	0.0292
13	0.0419	0.0382	0.0403
14	0.0536	0.0451	0.0535
15	0.0593	0.0618	0.0599
16	0.0430	0.0483	0.0498
17	0.0326	0.0421	0.0380
18	0.0622	0.0736	0.0532

Рисунок 3.4 – График изменения инфляции (Y_3) и ее оценок (M7 и M8), вычисленных по формулам (3.7 и 3.8) (стрелками отмечены экзаменационные точки)

Экспериментально были отобраны такие дополнительные аргументы: $X_{7,t} = X_{1,t-3}$; $X_{8,t} = X_{3,t-1}$; $X_{9,t} = X_{3,t-2}$; $X_{10,t} = X_{3,t-3}$; $X_{11,t} = X_{5,t-1}$. Длина выборки уменьшилась на 3 точки и составила 15 точек, из них три последние – экзаменационные. Получена модель M8 вида:

$$V_{3,t} = 0.0544 + 0.0006 \times X_{2,t} + 0.0149 \times X_{3,t} - 0.0002 \times X_{4,t} + 0.0001 \times X_{1,t-3} + \\ + 0.0027 \times X_{3,t-1} + 0.0007 \times X_{3,t-2} + 0.0032 \times X_{5,t-1}. \quad (3.8)$$

Модель имеет следующие характеристики:

$$СКО = 0.0034; \quad R = 0.0089; \quad S = 16.58\%.$$

В таблице 3.6 приведены оценки инфляции (M8) по модели (3.8), на рисунке 3.4 представлен график их изменения. На всей выборке модель имеет хорошую точность,

которую удалось существенно повысить за счет предоставляемых системой АСТРИД средств. При этом применение МГУА совместно с введением лаговых переменных позволило предсказать резкое изменение тенденции развития инфляционного процесса. Отметим, что такой результат получен по данным, относящимся к стадии монотонного течения процесса, т.е. изменение тенденции неявно содержалось в данных этой стадии, и благодаря МГУА эту закономерность удалось обнаружить.

3.2. Индуктивный компьютерный советчик для прогнозирования макроэкономики Украины

Прогнозирование процессов, происходящих в макроэкономике, носит нормативный характер и потому можно построить компьютерный советчика для использования в процессе принятия решений при текущем регулировании макроэкономики [33-35]. Нормативные прогнозы можно получать с помощью моделей, построенных по алгоритмам МГУА [36,37] в результате обработки наблюдений за работой системы в течение нескольких лет. Обычно достаточно располагать среднеквартальными значениями переменных. Данные должны охватывать интервал времени, в течение которого законы, действующие в системе, существенно не изменялись. Ниже рассмотрен пример построения моделей макроэкономики Украины по данным за 1992-1995 годы.

Варианты нормативных прогнозов зависят от предоставленной компьютеру априорной информацией. Особенности важно решение следующих вопросов:

- 1) какое полное множество переменных предоставляется компьютеру для синтеза и перебора прогнозирующих моделей по внешнему критерию. При выборе такого множества следует использовать опыт экономистов, знающих систему. Может помочь и перебор вариантов множества переменных: тот вариант лучше, при котором минимум критерия вариации ошибки прогноза RR_{min} [38] меньше.
- 2) какие переменные отнесены к подмножеству входных факторов, а какие - к подмножеству выходных оптимизируемых переменных. Входные факторы нормативного прогноза обычно выбираются среди регулирующих воздействий макроэкономической системы и служат аргументами прогнозирующих моделей. Модели и цели прогноза и управления различны и, следовательно, множества входных факторов и регулирующих воздействий могут не совпадать.
- 3) какие ограничения заданы на допустимую величину входных факторов и выходные переменные.

Ниже на примере Украины описан метод нормативного прогнозирования процессов в макроэкономике с целью ее регулирования.

Характеристика исходного материала и выбор множества переменных. Обозначим входные факторы нормативного прогноза через x_i ($1 \leq i \leq M$), выходные переменные через y_j ($1 \leq j \leq L$) и внешние возмущающие воздействия через z_k ($1 \leq k \leq K$). В [33,34] находим, что макроэкономические системы характеризуются значениями следующих основных переменных:

y_1 - реальный валовый внутренний продукт (ВВП) (в млрд. крб. 1991г.);

y_2 - инфляция потребительских цен (% изменения);

y_3 - бюджетный дефицит (в % от ВВП);

y_4 - безработица (тысяч чел.);

x_1 - увеличение денежной базы (в % от ВВП);

x_2 - число приватизированных предприятий;

x_3 - индекс потребительских цен;

x_4 - скорость оборота денег.

z_1 - курс карбованца по неторговым операциям (относительно доллара США).

Значения этих переменных за 13 кварталов 1992-1995гг. приведены в таблице 3.7.

Таблица 3.7 – Исходные данные

Год, квартал	Y1	Y2	Y3	Y4	X1	X2	X3	X4	Z1
1992									
Q1	86.4	3.0	24.1	13.2	13.2	5	75.4	65.1	160
Q2	107.6	8.1	35.6	11.1	11.1	10	18.1	100	200
Q3	97.5	9.3	60.7	18.5	18.5	20	19.2	123.3	293
Q4	96.4	17.7	70.5	45.2	45.2	30	28.5	81	788
1993									
Q1	84.7	1.1	79.5	32.4	32.4	430	39.7	54.3	1893
Q2	78.9	13.2	73.3	35.9	35.9	830	39.4	38.2	3133
Q3	72.3	6.2	78.7	27.7	27.7	1685	44.5	38.9	8350
Q4	56.4	6.1	83.9	12.2	12.2	3585	66.4	20.6	26667
1994									
Q1	38.8	6.4	98.6	11.8	11.8	5442	12.4	26.8	35597
Q2	42.4	8.9	92.8	12.9	12.9	8402	5.0	33.3	43466
Q3	46.9	19.8	88.9	22.6	22.6	10214	4.0	45.7	46657
Q4	44.5	8.1	82.2	7.3	7.3	11552	39.5	25.4	110242
1995									
Q1	36.2	8.0	86.6	3.7	3.7	12802	16.8	22.8	142782

Удобно привести данные к безразмерному виду с помощью нормирования:

$$Y_j = (y_j - y_{j \min}) / (y_{j \max} - y_{j \min}); \quad X_i = (x_i - x_{i \min}) / (x_{i \max} - x_{i \min}).$$

Таблица 3.8 – Нормированные исходные данные

Год и квартал	Y1	Y2	Y3	Y4	X1	X2	X3	X4	Z1
1992									
Q1	0.703	0.102	0.000	1.000	0.229	0.000	1.000	0.433	0.000
Q2	1.000	0.374	0.154	0.156	0.178	0.000	0.198	0.773	0.000
Q3	0.859	0.438	0.491	0.168	0.357	0.001	0.213	1.000	0.001
Q4	0.843	0.888	0.623	0.271	1.000	0.002	0.343	0.588	0.004
1993									
Q1	0.679	0.000	0.744	0.410	0.692	0.033	0.500	0.328	0.012

Q2	0.598	0.647	0.660	0.406	0.776	0.064	0.496	0.171	0.021
Q3	0.506	0.273	0.733	0.466	0.578	0.131	0.567	0.178	0.057
Q4	0.283	0.267	0.803	0.659	0.205	0.280	0.874	0.000	0.186
1994									
Q1	0.036	0.283	1.000	0.093	0.195	0.425	0.118	0.060	0.249
Q2	0.087	0.417	0.922	0.011	0.222	0.656	0.014	0.124	0.304
Q3	0.150	1.000	0.870	0.000	0.455	0.798	0.000	0.244	0.326
Q4	0.116	0.374	0.780	0.407	0.087	0.902	0.497	0.056	0.772
1995									
Q1	0.000	0.369	0.839	0.141	0.000	1.000	0.179	0.021	1.000
Q2									1.070

Разностные прогнозирующие оптимальные нефизические модели при явных шаблонах. Шаблоном называется граф, показывающий, какие запаздывающие аргументы предлагаются компьютеру на выбор (рисунок 3.5 а,б,с). При явных шаблонах (а) точность прогноза меньше, но вычисления проще. Применение неявных шаблонов (б, с) приводит к повышению точности, но связано с решением системы линейных уравнений для каждой точки нормативного прогноза.

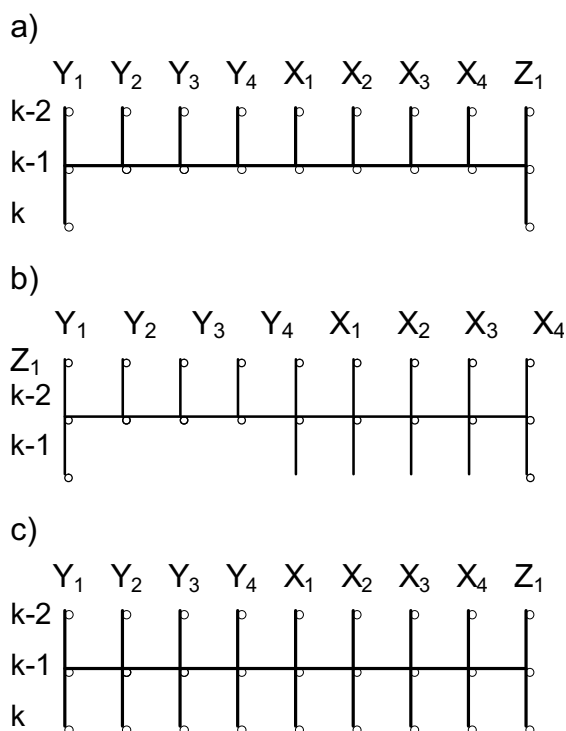


Рисунок 3.А – Шаблоны: а) явный для пошагового прогноза переменной Y_1 ; б) явный для нормативного прогноза Y_1 ; в) неявный для нормативного прогноза переменных $Y_{1,k}, Y_{2,k}, Y_{3,k}$ и $Y_{4,k}$.

Для прогноза первой переменной $Y_{1,k}$, пользуясь явным шаблоном (b) (рисунок 3.5), составим полное уравнение регрессии следующего вида:

$$Y_{1,k} = a_{00} + a_{11} \times Y_{1,k-1} + a_{12} \times Y_{1,k-2} + a_{21} \times Y_{2,k-1} + a_{22} \times Y_{2,k-2} + a_{31} \times Y_{3,k} + a_{32} \times Y_{3,k-2} + \\ + a_{41} \times Y_{4,k-1} + a_{42} \times Y_{4,k-2} + a_{50} \times X_{1,k} + a_{51} \times X_{1,k-1} + a_{52} \times X_{1,k-2} + a_{60} \times X_{2,k} + \\ + a_{61} \times X_{2,k-1} + a_{62} \times X_{2,k-2} + a_{70} \times X_{3,k} + a_{71} \times X_{3,k-1} + a_{72} \times X_{3,k-2} + a_{80} \times X_{4,k} + \\ + a_{81} \times X_{4,k-1} + a_{82} \times X_{4,k-2} + a_{90} \times Z_{1,k} + a_{91} \times Z_{1,k-1} + a_{92} \times Z_{1,k-2}.$$

Полные уравнения аналогичного вида составляются и для остальных выходных переменных $Y_{2,k}$, $Y_{3,k}$ и $Y_{4,k}$. Далее по комбинаторному алгоритму МГУА находятся значения коэффициентов этих уравнений, рассчитываются прогнозы и оценивается их точность.

Пример нормативного прогноза. Для примера рассмотрим нормативный прогноз макроэкономики Украины на второй квартал 1995 года по данным, приведенным в таблицах 3.7 и 3.8. Переменные с индексом k относятся к этому кварталу.

Для модели прогноза $Y_{1,k}$ получены следующие значения коэффициентов:

$a_{00} = -0.823$;	$a_{11} = 0$;	$a_{12} = 0$;	$a_{20} = 0$;	$a_{21} = -1.527$;	$a_{22} = -1.539$;
$a_{30} = 0$;	$a_{31} = 0.549$;	$a_{32} = 0$;	$a_{40} = 0$;	$a_{41} = -0.120$;	$a_{42} = 0.798$;
$a_{50} = 0$;	$a_{51} = 0$;	$a_{52} = 0.295$;	$a_{60} = 1.226$;	$a_{61} = -0.800$;	$a_{62} = 0$;
$a_{70} = 1.311$;	$a_{71} = 0$;	$a_{72} = -0.581$;	$a_{80} = 0.860$;	$a_{81} = 0$;	$a_{82} = 2.121$;
$a_{90} = 0.734$;	$a_{91} = -0.467$;	$a_{92} = 2.364$.			

Точность моделей характеризуется глубиной минимума критерия RR . Для оптимальной нефизической модели прогноза $Y_{1,k}$ получено $RR_{min} = 1.59e^{-14}$ (СКО=0.0). Для прогноза переменной $Y_{2,k}$ $RR_{min} = 7.55e^{-14}$ (СКО = $3.92e^{-15}$); для $Y_{3,k}$ $RR_{min} = 3.63e^{-5}$ (СКО = $6.77e^{-6}$); для $Y_{4,k}$ $RR_{min} = 0.0026$ (СКО = $1.05e^{-4}$).

Малая величина минимального значения критерия указывает на высокую точность прогнозов переменной $Y_{1,k}$. Аргументы, при которых коэффициент равен нулю, исключаются из модели – их не следует использовать для получения точного прогноза. Выводы о целесообразности управления объектом по прогнозирующей нефизической модели делать не следует. Для решения вопросов управления нужен анализ физической модели объекта, которая здесь не рассматривается.

Значения всех переменных в предыдущие моменты $k-1$ и $k-2$ известны. Подставив их в уравнения для переменных $Y_{1,k}$, $Y_{2,k}$, $Y_{3,k}$ и $Y_{4,k}$, для явных шаблонов получим четыре несвязанных расчетных уравнения:

$$Y_{1,k} = a_0 + a_1 \times X_{1,k} + a_2 \times X_{2,k} + a_3 \times X_{3,k} + a_4 \times X_{4,k} + a_5 \times Z_{1,k}, \quad (3.9)$$

$$Y_{2,k} = b_0 + b_1 \times X_{1,k} + b_2 \times X_{2,k} + b_3 \times X_{3,k} + b_4 \times X_{4,k} + b_5 \times Z_{1,k}, \quad (3.10)$$

$$Y_{3,k} = c_0 + c_1 \times X_{1,k} + c_2 \times X_{2,k} + c_3 \times X_{3,k} + c_4 \times X_{4,k} + c_5 \times Z_{1,k}, \quad (3.11)$$

$$Y_{4,k} = d_0 + d_1 \times X_{1,k} + d_2 \times X_{2,k} + d_3 \times X_{3,k} + d_4 \times X_{4,k} + d_5 \times Z_{1,k}. \quad (3.12)$$

Для уравнения (3.9) находим:

$$a_0 = -2.213; \quad a_1 = 0 \quad a_2 = 1.226; \quad a_3 = 1.311; \quad a_4 = 0.860; \quad a_5 = 0.734.$$

Индуктивный компьютерный алгоритм показывает, что для получения точного нормативного прогноза переменной $Y_{1,k}$ нужно задать значения $X_{2,k}$, $X_{3,k}$ и $X_{4,k}$.

Для уравнения (3.9) получим:

$$b_0 = 0.337; b_1 = 0.564; b_2 = 0; b_3 = 0; b_4 = 1.542; b_5 = 0.$$

Для нормативного прогноза переменной $Y_{2,k}$ следует задать факторы $X_{1,k}$ и $X_{4,k}$. Для уравнения (3.11) получим:

$$c_0 = -0.639; c_1 = 0.088; c_2 = 1.477; c_3 = 0; c_4 = 0; c_5 = 0.$$

Для нормативного прогноза переменной $Y_{3,k}$ следует задать факторы $X_{1,k}$ и $X_{2,k}$. Наконец, для уравнения (3.12) находим

$$d_0 = 0.003; d_1 = 0; d_2 = 0; d_3 = 0.756; d_4 = 0; d_5 = 0.$$

Для наиболее точного нормативного прогноза переменной $Y_{4,k}$ следует задать значение только одного фактора $X_{3,k}$.

Графическая визуализация нормативного прогноза переменной $Y_{1,k}$. Покажем визуализацию нормативного прогноза выходной переменной $Y_{1,k}$ на рассмотренном выше примере. Зададим переменной $X_{4,k}$ два значения: $X_{4,k} = 0.25$ и $X_{4,k} = 0.75$.

Изменение переменной $Y_{1,k}$ показано при помощи изолиний на плоскости двух входных факторов $X_{2,k}$ и $X_{3,k}$ (рисунок 3.6, а и б). Сравнивая рисунки 3.6 а и б, можно, в частности, заключить, что повышение скорости оборота денег приводит к увеличению ВВП. Главное в том, что по рисунку можно установить количественную связь между входными факторами и выходной переменными.

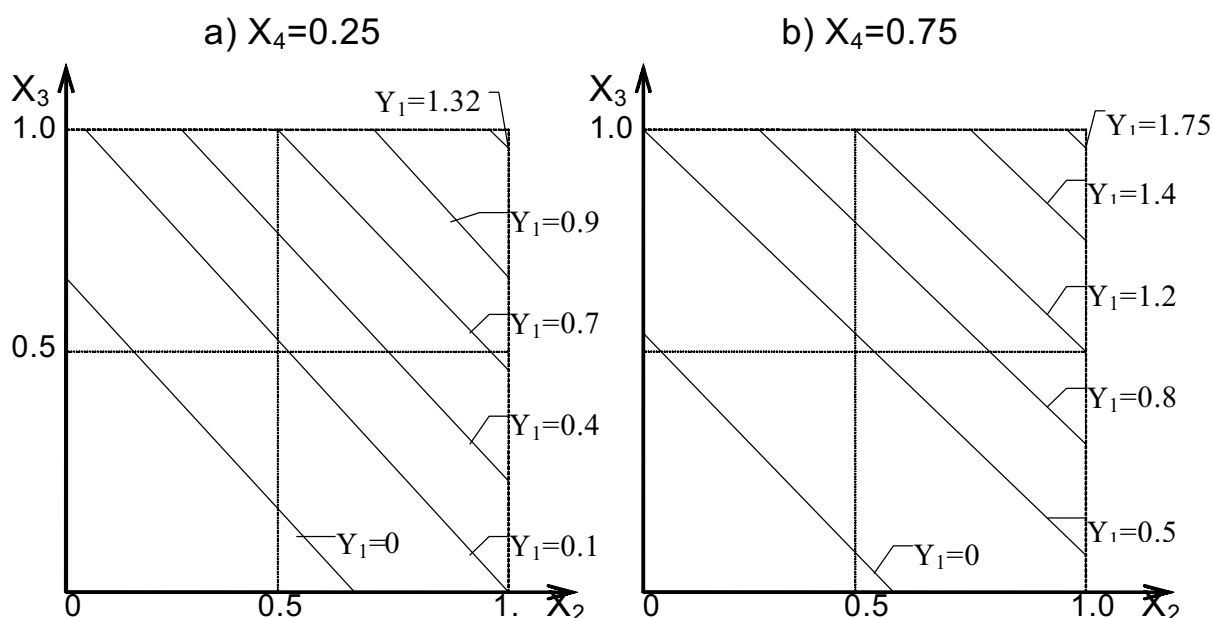


Рисунок 3.В – Графическое изображение нормативного прогноза $Y_{1,k}$

Отдельный прогноз возмущающего воздействия. Вспомогательный прогноз внешнего возмущения $Z_{l,k}$ выполняется отдельно по комбинаторному алгоритму МГУА в виде функции запаздывающих значений переменных, указанных в таблице 3.8. Полное уравнение, предлагаемое компьютеру, имеет следующий вид:

$$\begin{aligned} Z_{1,k} = & e_0 + e_1 \times Z_{1,k-1} + e_2 \times Z_{1,k-2} + e_3 \times Z_{1,k-3} + e_4 \times Z_{1,k-4} + e_6 \times Y_{1,k-1} + e_7 \times Y_{1,k-2} + \\ & + e_8 \times Y_{1,k-3} + e_9 \times Y_{1,k-4} + e_{11} \times Y_{2,k-1} + e_{12} \times Y_{2,k-2} + e_{13} \times Y_{2,k-3} + e_{14} \times Y_{2,k-4} + \\ & + e_{16} \times Y_{3,k-1} + e_{17} \times Y_{3,k-2} + e_{18} \times Y_{3,k-3} + e_{19} \times Y_{3,k-4} + e_{21} \times Y_{4,k-1} + e_{22} \times Y_{4,k-2} + \\ & + e_{23} \times Y_{4,k-3} + e_{24} \times Y_{4,k-4} + e_{26} \times X_{1,k-1} + e_{27} \times X_{1,k-2} + e_{28} \times X_{1,k-3} + e_{29} \times X_{1,k-4} + \\ & + e_{31} \times X_{2,k-1} + e_{32} \times X_{2,k-2} + e_{33} \times X_{2,k-3} + e_{34} \times X_{2,k-4} + e_{36} \times X_{3,k-1} + e_{37} \times X_{3,k-2} + \\ & + e_{38} \times X_{3,k-3} + e_{39} \times X_{3,k-4} + e_{41} \times X_{4,k-1} + e_{42} \times X_{4,k-2} + e_{43} \times X_{4,k-3} + e_{44} \times X_{4,k-4} \end{aligned}$$

В результате применения предварительной селекции простых линейных моделей вида $y = a_0 + a_i X_i$ из 36-ти указанных аргументов выбираем $F=20$ наиболее значимых:

$$\begin{array}{cccccccccccc} Z_{1,k-1} & X_{2,k-2} & X_{2,k-3} & Z_{1,k-2} & Y_{1,k-2} & Y_{1,k-1} & Y_{3,k-2} & X_{2,k-4} & X_{2,k-1} & Y_{1,k-3} \\ X_{1,k-1} & Y_{1,k-4} & X_{4,k-4} & X_{4,k-4} & X_{1,k-2} & Y_{3,k-1} & Y_{4,k-1} & X_{3,k-1} & Y_{2,k-1} & X_{4,k-2} \end{array}$$

Для этого множества факторов по комбинаторному алгоритму получена прогнозирующая модель:

$$Z_{1,k} = 0,61 - 0,16Y_{2,k-1} - 0,06X_{3,k-1} - 0,34X_{3,k-3} - 0,46X_{4,k-4} + 2,19Z_{1,k-4}.$$

Критерий вариации ошибки прогноза $RR_{min} = 0.00028$. Прогноз на второй квартал 1995г. равен: $Z_{l,k} = 1.07$, и это значение было использовано выше для прогноза выходных переменных.

3.2. Сравнение прогнозирующих свойств моделей, полученных с применением регрессионного анализа и МГУА

Сравним прогнозирующие свойства моделей МНК и МГУА на двух примерах.

Моделирование объема производства легкой промышленности. Исходные данные приведены в таблице 3.3 (см. 3.3). В этом примере для проверки прогнозных свойств модели, оставим последние 5 экзаменационных точек. Это связано с тем, что для оценки 14 коэффициентов (13 аргументов и свободный член) по МНК и для вычисления их статистических оценок требуется по крайней мере 15 точек.

Модель МНК, включающая все тринадцать исходных аргументов, имеет вид (М9):

$$\begin{aligned} V_t = & 1.0849 + 0.0315 \times X_{1,t} + 0.0015 \times X_{2,t} + 0.0683 \times X_{3,t} + 0.0015 \times X_{4,t} - 0.0144 \times X_{5,t} + \\ & + 0.1437 \times X_{6,t} - 0.0064 \times X_{7,t} + 0.0977 \times X_{8,t} - 0.0046 \times X_{9,t} - 0.0457 \times X_{10,t} - \\ & - 0.0111 \times X_{11,t} + 0.0069 \times X_{12,t} - 0.0019 \times X_{13,t}; \end{aligned} \quad (3.13)$$

$$\text{СКО} = 0.03168; R = 0.1102; S = 118.5\%.$$

Статистические оценки модели: $R^2 = 0.984$, $R_y^2 = 0.782$, где R^2 - коэффициент детерминации, R_y^2 - скорректированный выборочный коэффициент детерминации [39].

Значения статистики t_i ($i = 0, \dots, 13$) при табличном $t_{(0.05, 15)} = 1.753$:

$$\begin{aligned} t_0 &= 0.388; t_1 = 0.631; t_2 = 0.228, t_3 = 1.948; t_4 = 0.676; t_5 = 0.681, t_6 = 0.502; \\ t_7 &= 0.262; t_8 = 0.784, t_9 = 0.307; t_{10} = 0.339, t_{11} = 0.666; t_{12} = 0.657; t_{13} = 0.386. \end{aligned}$$

В таблице 3.9 приведены оценки объема производства легкой промышленности (МНК), полученные по модели М9(3.13), на рисунке 3.6 представлен график их изменения. Видно, что модель довольно плохая, причем наибольшие ошибки относятся к экзаменационным точкам. Только один коэффициент оказался значимым ($t_3 > 1.753$), но приравнять нулю остальные коэффициенты без дополнительных исследований нельзя [39]. То есть МНК не позволяет построить модель, отражающую системные закономерности и пригодную для получения надежного прогноза, несмотря на то, что некоторые статистические оценки модели оказались значимыми.

Модель МГУА (М10), построенная при тех же условиях, имеет вид:

$$V_{2,t} = -0.1425 + 0.0046 \times X_{2,t} + 0.0491 \times X_{3,t} + 0.0022 \times X_{4,t} - 0.0826 \times X_{6,t} + 0.0044 \times X_{9,t}. \quad (3.14)$$

Статистические оценки модели:

$$R^2 = 0.958; R_y^2 = 0.934; F = 40.82; m1 = 5; m2 = 9; F_{(0.05, 5, 9)} = 3.48,$$

где F , $F_{(\alpha, m1, m2)}$ - расчетное и табличное значения критерия Фишера с уровнем значимости α и степенями свободы $m1$ и $m2$.

Значения статистики t_i ($i = 0, 2, 3, 4, 6, 9$) при табличном $t_{(0.05, 15)} = 1.753$:

$$t_0 = 43.434; t_2 = 5.899; t_3 = 6.132; t_4 = 5.216; t_6 = 3.573; t_9 = 2.769.$$

В таблице 3.9 приведены оценки объема производства по модели М10(3.14), на рисунке 3.6 – график их изменения. Видно, что прогнозирующие свойства модели МГУА значительно выше, и она имеет все значимые коэффициенты.

Таблица 3.F – Значения оценок выходной величины

н/п	Y2	V2	
		M9	M10
1	0.1623	0.1632	0.1604
2	0.1468	0.1487	0.1524
3	0.1106	0.1108	0.1118
4	0.1279	0.1225	0.1267
5	0.1378	0.1404	0.1389
6	0.1417	0.1402	0.1389
7	0.1096	0.1135	0.1114
8	0.0995	0.0979	0.0976
9	0.0938	0.0943	0.0991
10	0.0979	0.1025	0.0969

н/п	Y2	V2	
		M9	M10
11	0.1022	0.0995	0.0927
12	0.1200	0.1155	0.1159
13	0.1065	0.1037	0.0997
14	0.1093	0.1108	0.1144
15	0.0693	0.0718	0.0782
16	0.0917	0.1056	0.0870
17	0.0964	0.1541	0.1157
18	0.0941	0.2043	0.0932
19	0.0882	0.0733	0.0914
20	0.0866	0.0313	0.0905

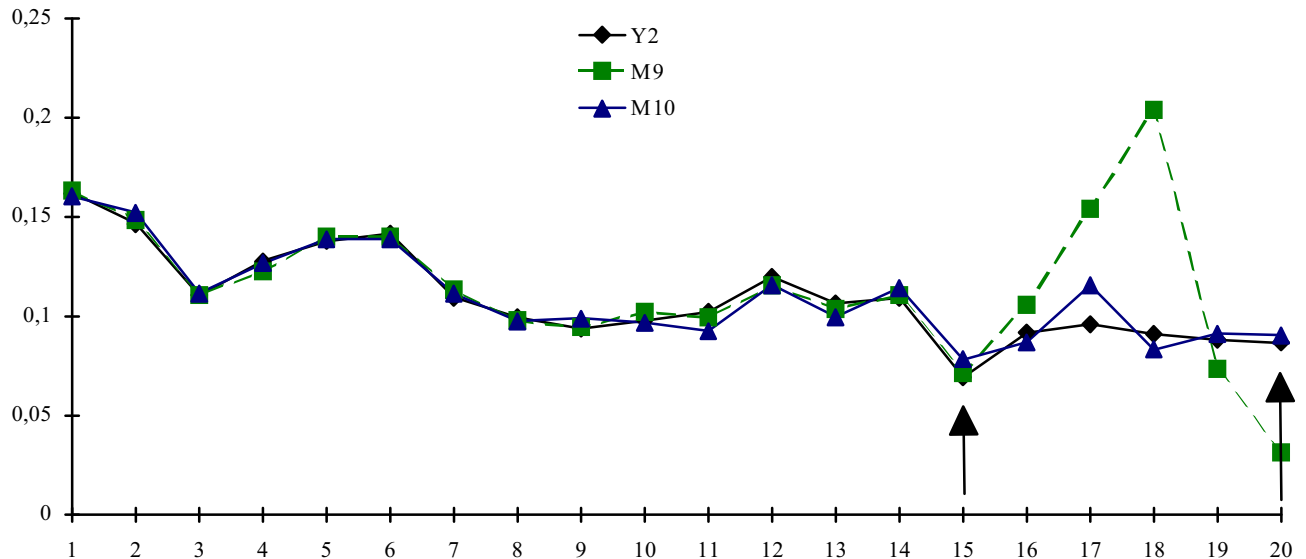


Рисунок 3.6 – Сравнение моделей объема производства легкой промышленности, полученных по МНК (M9) и по МГУА (M10) (стрелками отмечены начало и конец экзаменационной последовательности)

Моделирование инфляции. Исходные данные приведены в таблице 3.5 (см. 3.4). Строились модели зависимости инфляции от текущих значений аргументов, причем 3 последние точки составляли экзаменационную последовательность, т.е. для получения модели использовались только первые 15 точек. Выбор такого варианта расчетов связан с резким изменением характера развития процесса. Ставилась задача: можно ли, используя данные, относящиеся к периоду монотонного развития инфляции, предсказать это резкое изменение? Другими словами, следовало проверить, заложено ли это изменение в предыстории и можно ли его выявить с помощью моделирования.

Модель МНК, включающая все шесть исходных аргументов, имеет вид:

$$V_{3,t} = 0.1646 + 0.00015 \times X_{1,t} - 0.00702 \times X_{2,t} - 0.00996 \times X_{3,t} + 0.00143 \times X_{4,t} + 0.00031 \times X_{5,t} - 0.00056 \times X_{6,t}, \quad (3.15)$$

где t – номер точки, V_3 – модельная оценка выходной переменной Y_3 , и имеет следующие точностные и статистические характеристики:

$$\text{СКО} = 0.0297; \quad R = 0.0854; \quad S = 157.9\%;$$

$$R^2 = 0.864; \quad R_y^2 = 0.762; \quad F = 8.475; \quad m_1 = 6; \quad m_2 = 8; \quad F_{(0.05,6,8)} = 3.58.$$

Значения статистики t_i ($i = 0, 1, \dots, 6$) при табличном $t_{(0.05,15)} = 1.753$:

$$t_0 = 2.9155; \quad t_1 = 0.296; \quad t_2 = 1.619; \quad t_3 = 1.005; \quad t_4 = 0.685; \quad t_5 = 1.801; \quad t_6 = 2.927;.$$

В таблице 3.10 приведены оценки инфляции по модели МНК (3.15), на рисунке 3.7 представлен график их изменения. Видно, что модель довольно плохая, причем наибольшие ошибки относятся к трем экзаменационным точкам, когда тенденция развития инфляции резко изменилась. Таким образом, МНК не позволяет построить модель, отражающую системные закономерности и пригодную для получения надежного прогноза,

несмотря на то, что модель по статистическим оценкам значима: $F > F_{(0.05,6,8)}$; $t_0, t_5, t_6 > t_{(0.05,15)}$.

Модель, полученная при тех же условиях по МГУА (М7), имеет вид (см. (3.7)):

$$V_{3,t} = 0.00036 \times X_{1,t} + 0.0109 \times X_{3,t} - 0.00046 \times X_{4,t} + 0.00019 \times X_{5,t} - 0.0001156 \times X_{6,t}$$

и имеет такие показатели качества и статистические характеристики:

$$\text{СКО} = 0.00873; R = 0.0194; S = 35.9\% ;$$

$$R^2 = 0.718; R_y^2 = 0.604; F = 6.796; F_{(0.005,4,10)} = 3.48;$$

Значения статистики t_i ($i = 1,3,4,5,6$) при табличном $t_{(0.05,15)} = 1.753$:

$$t_1 = 0.631; t_3 = 1.252; t_4 = 0.816; t_5 = 1.299; t_6 = 2.646.$$

Качество этой модели наглядно характеризует также рисунок 3.7. Видно, что она четко отражает изменение тенденции процесса, не очевидное из предыдущей информации, т.е. из предыстории (до 16-й точки) процесса.

Важно отметить, что улучшение прогноза состоялось за счет *упрощения прогнозирующей модели* (исключения из нее аргумента X_2), что характерно именно для применения МГУА (эффект исключения “лишних”, неинформативных факторов).

Таблица 3.G — Оценки процесса инфляции

п/н	Y3	V3
		M11
1	0.0145	0.0125
2	0.0357	0.0163
3	0.0267	0.0107
4	0.0081	0.0187
5	0.0160	0.0205
6	0.0092	0.0174
7	0.0117	0.0147
8	0.0116	0.0152
9	0.0140	0.0192
п/н	Y3	V3
		M11
10	0.0175	0.0227
11	0.0283	0.0307
12	0.0287	0.0305
13	0.0419	0.0382
14	0.0536	0.0451
15	0.0593	0.0618
16	0.0430	0.0483
17	0.0326	0.0421
18	0.0622	0.0736

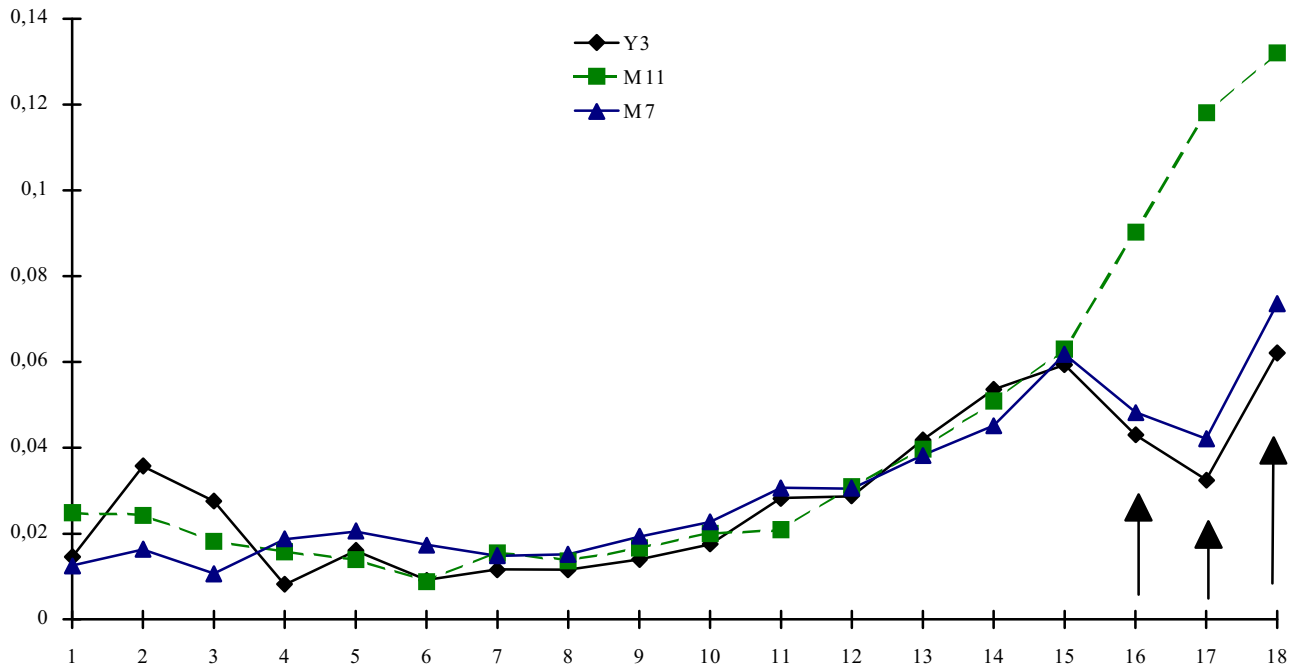


Рисунок 3.7 – Сравнение моделей инфляции (Y_3) и ее оценок (M7 и M11), вычисленных по формулам (3.7 и 3.15) (стрелками отмечены точки экзаменационной последовательности)

РАЗДЕЛ 4

РАЗРАБОТКА ОСНОВ МЕТОДОЛОГИИ СОЗДАНИЯ ИНТЕРАКТИВНЫХ СИСТЕМ МОДЕЛИРОВАНИЯ ПО ДАННЫМ НАБЛЮДЕНИЙ

4.1 Методика структуризации знаний в области автоматизированного построения моделей

При разработке современной программной системы структурной идентификации моделей, или моделирования по экспериментальным данным, возникает необходимость определенной структуризации знаний в данной области. Отметим, что задача структуризации знаний характерна для проблем разработки баз знаний экспертных систем, однако здесь структуризация применяется с сугубо прагматической целью - выработать систематизированные конструктивные рекомендации по разработке функционального и диалогового обеспечения программной системы моделирования.

Для структуризации знаний в рассматриваемой предметной области целесообразно проанализировать следующие аспекты проблемы:

- 1) определение основных этапов процесса решения задачи моделирования;
- 2) выделение основных методов решения такой задачи и сравнение их функциональных особенностей;
- 3) сравнительный анализ эффективности этих методов;
- 4) анализ и обобщение опыта их применения с формированием необходимых рекомендаций. Рассмотрим несколько подробнее указанные аспекты проблемы структуризации.

1. Основные этапы решения задачи моделирования определим, опираясь на ее формулировку в виде задачи дискретного (целочисленного) программирования (1.1). Ее анализ показывает, что в общем случае процесс решения этой задачи структурно-параметрической идентификации включает в себя следующие основные этапы:

- 1) Задание выборки данных (полученной в результате пассивного или активного эксперимента), а также априорной информации;
- 2) Выбор или задание класса базисных функций и соответствующее преобразование данных;
- 3) Генерация различных структур моделей в выбранном классе;
- 4) Оценивание параметров генерируемых структур и формирование множества F ;
- 5) Минимизация заданного критерия $CR(f)$ и выбор оптимальной модели f^* .
- 6) Проверка адекватности полученной оптимальной модели.
- 7) Принятие решения о завершении процесса.

Перечисленные этапы описывают произвольный процесс построения моделей, причем в зависимости от априорной информации и цели моделирования те или иные этапы могут отсутствовать. Например, в частном случае задачи параметрической идентификации множество F состоит из одной модели ("генерируется" единственная структура), т.е. исключается этап 5 процесса. Очевидно, что сформулированные этапы должны быть

положены в основу функциональной архитектуры программной системы структурной идентификации.

2. Пусть каким-либо образом (с помощью анализа литературы или опроса экспертов) отобрано некоторое множество методов структурной идентификации. Каждый из них решает задачу (1.1), и соответственно может быть охарактеризован такими основными компонентами:

- 1) *классом моделей* (базисных функций), для которого он предназначен;
- 2) применяемыми в нем *генератором структур*;
- 3) *методом оценивания параметров*;
- 4) *критерием качества моделей*.

Способы анализа адекватности моделей, сводящиеся к проверке правильности принятых априорных предположений, в основные характеристики метода могут не включаться.

Таким образом, каждый метод моделирования можно описать с помощью четырех указанных компонентов, с учетом которых легко можно сравнить функциональные особенности заданного множества методов. С другой стороны, для отобранной совокупности методов, расчленив каждый из них на такие составляющие, можно образовать *множества* классов моделей, генераторов структур, методов оценивания параметров и критериев выбора моделей и анализировать эти множества как самостоятельные научные объекты. Отметим, что эти четыре множества имеют тот же смысл, что и множества возможных вариантов выбора на этапах 2-5 процесса моделирования, т.е. указанное расчленение методов дает оригинальный ключ к формированию функционального наполнения программной системы в виде так называемых обобщенных функциональных модулей (см. ниже).

3. Сравнение эффективности различных методов моделирования сводится прежде всего к сравнению применяемых в них критериев выбора моделей. Это следует из того, что для конкретного моделируемого объекта при заданном объеме априорной информации можно, как правило, выбрать наиболее целесообразный вариант класса моделей, генератора структур и метода оценивания параметров, и тогда эффективность достижения желаемой цели моделирования определяется критерием качества моделей.

4. Анализ опыта применения алгоритмов необходим для построения правил выбора решений на каждом этапе моделирования, на основе которых разрабатывается структура диалога. От уровня обоснованности таких правил и их разнообразия по отношению к условиям моделирования зависит интеллектуальный уровень диалога и соответственно системы в целом. Кроме того, структура диалога в современной системе должна учитывать также возможный уровень квалификации пользователя. Для обеспечения подстройки системы к пользователю в работе предложен способ организации многоуровневого диалога.

При этом принято, что предполагаемую подготовленность можно отнести к одному из трех основных уровней:

- 5) *низкий* - начинающий пользователь (знакомый со своим объектом моделирования);
- 6) *средний* - подготовленный пользователь (знающий объект и общие принципы моделирования);
- 7) *высокий* - квалифицированный пользователь (эксперт, знающий и объект, и методы моделирования).

Указанным уровням квалификации ставится в соответствие три основных *режима диалога*: 1) *автоматический*, 2) *интерактивный*, 3) *меню-подобный*, сущность которых состоит в следующем. Для каждого этапа выбора решения составляется полное меню возможных решений (включая и отказ от выбора). Затем для автоматического режима фиксируется (по умолчанию) одно наиболее часто принимаемое решение; для интерактивного режима несколько наиболее часто (по опыту) выбираемых решений снабжаются подсказками и комментариями об условиях их предпочтения; в третьем режиме предлагается выбрать решение из полного меню без комментариев.

Поскольку такой принцип диалога применяется на каждом этапе моделирования, при работе системы возможен как один из трех основных режимов, так и произвольное их сочетание, изменяющееся от одного этапа к другому. Это дает эффект гибкой адаптации системы к уровню квалификации пользователя, т.е. к его информированности о процессе моделирования, к знанию им объекта и к объему имеющейся априорной информации. Более подробно вопросы построения диалога рассмотрены в следующем подразделе.

Итак, выше изложена *методика структуризации знаний* для разработки эффективной программной системы моделирования.

Эффективная в прикладном смысле система структурной идентификации должна характеризоваться: достаточной общностью в смысле использования основных методов идентификации; адаптируемостью к различным условиям априорной неопределенности; достаточной доступностью пользователям разной квалификации; вычислительной эффективностью (быстродействием). Как следует из предыдущего анализа, для достижения этих характеристик следует выполнить такие основные этапы как составляющие новой *методологии разработки интерактивных систем моделирования* с элементами искусственного интеллекта:

- различные методы решения такой задачи расчленяются на рассмотренные выше основные компоненты, совокупности которых анализируются независимо;
- разрабатываются обобщенные модули классов моделей, генераторов структур и критериев селекции;
- проектируется функциональная структура системы в соответствии с основными этапами моделирования;
- создаются процедуры диалога, способствующие достижению приемлемого результата пользователями различной квалификации.

4.2 Конструирование обобщенных многофункциональных вычислительных модулей

При разработке функциональной структуры современной проблемно-ориентированной программной системы, направленной на объединение ряда наиболее эффективных методов решения некоторого класса задач, обычно исходят из двух основных принципов [40,41]:

1) формирование библиотеки программ, реализующих отобранные методы. При этом все программы независимы, обладают собственными диалоговыми средствами и объединены специальным внешним монитором. Такой подход эффективен при наличии готовых качественных программ для каждого метода.

2) формирование библиотеки элементарных вычислительных модулей (процедур) с заданными правилами компоновки из этих модулей программ, реализующих отобранные методы. При этом все программы формируются головным монитором, который содержит также все необходимые для использования методов диалоговые средства. Такой подход эффективен при разработке всех программ "с нуля", на основе только известных из литературы математических описаний методов.

Как было показано в предыдущем подразделе, возможен и третий принцип формирования функциональной структуры программной системы, основанный на соответствующей структуризации знаний по конкретной проблеме. В результате структуризации определяются принципы формирования алгоритмических модулей, обобщающих функции ряда методов решения конкретного класса задач в единый многофункциональный алгоритм, характеризующийся определенным набором ключевых параметров, те или иные комбинации которых дают отдельные методы как частные случаи. Эффективность этого подхода определяется уменьшением числа и объема программных модулей, а также упрощением создания управляющего монитора.

Рассмотрим, как формируются структуры обобщенных модулей для выделенных выше трех основных компонент произвольного метода структурной идентификации на примере МГУА, опираясь на результаты соответствующей структуризации знаний [42].

Обобщенный генератор структур моделей. Комбинаторный алгоритм, вида (1.55), может быть сделан структурно гибким введением ограничения на минимальную SI и максимальную SA сложность генерируемых моделей. Тогда при комбинаторном переборе рассматриваются все модели сложности $s = \overline{SI, SA}$, и полному перебору соответствуют $SI=1$, $SA=m$. Если класс моделей полиномиальный, структура полного полинома (1.47) определяется степенью ST . При $SI = SA = m$ получаем одну модель, оцениваемую по МНК. Если при последовательном генерировании моделей сложности s учитывать не все C_m^{s-1} моделей, а только FS , улучшающих критерий, получим селекционно-комбинаторный алгоритм типа (1.56) как более общий по сравнению с комбинаторным – последний получим при максимальном $FS = C_m^{s-1}$.

Наконец, в случае $SI = SA = 2$ и максимальном FS получим перебор всех частных моделей от двух аргументов, т.е. первый ряд линейного итерационного алгоритма (1.57)

селекционного типа, а при $ST=2$ - также и нелинейного алгоритма (1.60). Следовательно, итерационный алгоритм с числом итераций NI может быть основан на гибком комбинаторном алгоритме. При этом возможности такого алгоритма автоматически расширяются: при $SI=SA>2$ получим "групповой учет" по три, четыре и т.д. аргументов, а при $SI=1, SA \geq 2$ - дополнительный полный перебор структуры частного описания. При $NI=1$ получаем комбинаторный, или однорядный алгоритм.

Все предыдущие алгоритмы представляют собой иерархически упорядоченную классификацию основных генераторов структур одномерных моделей ($MY=1$). Дальнейшее обобщение этих алгоритмов дает переход в область многомерных моделей с заданным числом выходных переменных $MY>1$. При идентификации таких моделей может быть использован любой из "одномерных" генераторов структур.

Итак, в предложенной классификации основных алгоритмов МГУА выделены ключевые параметры, определяющие: n, m - сложность задачи оценивания; $SI, SA, (ST), FS, NI$ - тип алгоритма генерации структур одномерных моделей; MY - вид решаемой многомерной задачи. При этом обобщенный генератор структур является модифицированным комбинаторным алгоритмом, в котором: вместо двоичного счетчика применяется последовательное формирование C_m^s структурных векторов для каждого $s = \overline{1, m}$; вводятся ограничения SI, SA на сложность рассматриваемых моделей, $1 \leq SI \leq s \leq SA \leq m$, а также на число лучших моделей FS , допускаемых к дальнейшему усложнению, $FS_s \leq FM \leq C_n^s$; добавляется возможность итерационного (циклического) использования алгоритма NI раз.

Тогда *единый модуль с ключевыми параметрами $\{SI, SA, FS, NI\}$* обобщает ряд типов генераторов структур (относящихся не только к МГУА): многомерный регрессионный анализ ($SI=SA=M$), полный (комбинаторный) перебор (алгоритм всех регрессий) ($SI=1, SA=m, FS=C_m^s$); регрессионная процедура "включения" ($FS=1$); селекционно-комбинаторный генератор ($1 < FS < C_m^s$); многорядная (итерационная) процедура МГУА ($SI=SA=2, NI>1$), и т.д.

Обобщение классов моделей, линейных по параметрам. В задачах моделирования статических и динамических объектов и временных рядов применяются прежде всего полиномы (1.47) и разностные уравнения (1.53), которые вместе относятся к широкому классу моделей, линейных по параметрам. Рассмотрим возможные пути обобщения и универсализации указанных классов моделей (1.47), (1.53).

Возможности формализованного применения разнообразных нелинейных функций значительно расширяются при введении в (1.47) одночленов следующего весьма общего вида:

$$\prod_{v=1}^r X_{\alpha v}^{Z(c_{jv})^P}, \quad (4.1)$$

где r - число входных переменных (аргументов), одновременно учитываемых в мультипликативном члене (4.1), $1 \leq CI \leq r \leq CA \leq m$; α - индекс (номер) сочетания по r аргументов из m , $\alpha = \overline{1, C_m^R}$; Z, P - величины, принимающие значения $+1, 0, -1$; c_{jv} могут быть равными $1, 2, \dots, ST$ при соблюдении ограничения типа (1.48) с r вместо m и $c = \overline{r, ST}$. Тогда изменением чисел CI, CA, Z, P можем получить различные варианты нелинейных базисных функций: если $Z = P = 1$ то при $CI = 1, CA = m$ получим (1.7), при $CI = CA = 1$ - полином без произведений аргументов, при $CI > 1$ - полином только с мультипликативными членами, и т.д.; если $Z = 0$, получим свободный член полинома (1.7); при $Z = 1, P = 0$ все аргументы будут иметь единичные степени; при $Z = 1, P = -1$ степени будут дробными; при $Z = -1, P = 1$ получим обратные значения аргументов в различных степенях, и т.д. Таким образом, единая процедура формирования одночленов вида (4.1) позволяет с помощью изменения шестерки чисел $\{m, ST, CI, CA, Z, P\}$ автоматически генерировать разнообразные базисные функции полиномиального типа.

Что касается модели вида (1.53), то она обобщает ряд классов линейных моделей. Рассмотрим это несколько подробнее. Например, (1.35) превращается в обычную авторегрессионную модель с QY запаздываниями:

$$y_k = a_o + \sum_{j=1}^{QY} a_j y_{k-j}, \quad (4.2)$$

которая при учете MW внешних воздействий обобщается в модель с лаговыми (запаздывающими на QW тактов) переменными:

$$y_k = a_o + \sum_{j=1}^{QY} a_j y_{k-j} + \sum_{i=1}^{MW} \sum_{j=1}^{QW} b_{ij} w_{k-j+1}^i. \quad (4.3)$$

При многомерном выходе $MY > 1$ получаем обыкновенную векторную разностную модель ($l = \overline{1, MY}$):

$$y_k^l = a_{ot}^l + \sum_{i=1}^{MY} \sum_{j=1}^{QY} a_{ij}^l y_{k-j}^i + \sum_{i=1}^{MW} \sum_{j=1}^{QW} b_{ij}^l w_{k-j+1}^i. \quad (4.4)$$

В уравнения (4.1)-(4.4) могут быть добавлены члены в виде скользящих средних значений как зависимых y_k^l , так и независимых w_k^l величин. Очевидно также, что при $QY = 0, QW = 1$ получим систему регрессионных уравнений, описывающих статический объект:

$$y^l = a_o^l + \sum_{i=1}^{MW} b_i^l w^i, \quad l = \overline{1, MY}, \quad (4.5)$$

т.е. при различных сочетаниях чисел MY, QY, MW, QW можно получить классы моделей (4.1)-(4.5) и их разновидности.

Таким образом, мы фактически получили удобную классификацию линейных моделей, охватывающую большинство встречающихся на практике вариантов одномерных и

многомерных линейных моделей, или обобщенный многофункциональный модуль классов структур линейных моделей, который, например, по отношению к уравнению (4.39) характеризуется *ключевыми параметрами* $\{MY, QY, MW, QW\}$. Причем эта классификация, пример которой приведен в табл.5.2, удобна для программирования и реализована в описанном выше алгоритме ARIMAD.

Таблица 4.1 – Основные классы линейных моделей (частные случаи (4.4))

Тип по MY	Класс модели	MY	QY	MW	QW
Одно-мерные	регрессия	1	0	m	1
	авторегрессия (AP)	1	>0	0	0
	запаздывания по вх.(CC)	1	0	>0	>1
	APCC	1	>0	>0	>1
Много-мерные	многомерная регрессия	>1	0	m	1
	многомерная APCC	>1	>0	>0	>1

Обобщение критериев качества моделей. Такое обобщение можно рассмотреть на основе приведенных в главе 1 формул (1.7) - (1.9) для критериев селекции моделей, формирующихся без разделения выборки, и формул (1.29) - (1.33) для критериев, основанных на разбиении выборки, а также привести более общую формулу.

Рассмотренные в главе 1 наиболее известные критерии (1.7) - (1.9), основанные на вычислении RSS , содержат аддитивные или мультипликативные корректирующие добавки, явно зависящие от сложности s модели (числа оцениваемых параметров). Такая корректировка является “штрафом” за сложность модели и компенсирует необоснованное уменьшение RSS при возрастании s , обеспечивая тем самым минимум критерия $CR(s)$, указывающий “оптимальный” состав регрессоров модели как решение задачи структурной идентификации. Минимумы различных критериев, вообще говоря, не совпадают.

Нетрудно убедиться, что критерии (1.7) - (1.9) допускают единое представление в обобщенном виде

$$CR(s) = \eta_1(n, s)RSS(s) + \eta_2(n, s)\hat{\sigma}^2, \quad (4.6)$$

где $\eta_1(s, n)$, $\eta_2(s, n)$ - мультипликативный и аддитивный корректирующие члены (функции штрафа за сложность модели), $\hat{\sigma}^2$ - та или иная оценка неизвестной дисперсии σ^2 .

Очевидно, что внешние критерии МГУА (1.29) - (1.33), вычисляемые на двух подвыборках, можно также записать в общем виде через величины типа (1.27):

$$KR = \alpha_1 \Delta(W|A) + \alpha_2 \Delta(W|B) + \alpha_3 \varepsilon_W + \alpha_4 \varepsilon_A + \alpha_5 \varepsilon_B, \quad (4.7)$$

где $\alpha_1, \dots, \alpha_5$ принимают значения $\{1, 0, -1\}$. Теперь эти критерии можно включить в классификационную формулу (4.6), несколько обобщив ее:

$$CR(s) = \eta_1(n, s)V(s) + \eta_2(n, s)\hat{\sigma}^2. \quad (4.8)$$

где $V(s)$ - некоторый показатель качества модели. Тогда при $V(s)=RSS(s)$ получаем (4.6), а при $\eta_1(s, n)=1$, $\eta_2(s, n)=0$, $V(s)=KR(s)$ получаем (4.7) и соответствующие частные случаи. Таким образом, *ключевые параметры* $\{\eta_1(\cdot), \eta_2(\cdot), V(s), \hat{\sigma}^2\}$ характеризуют обобщенный многофункциональный модуль критериев, применяемых на практике в задаче структурной

идентификации прогнозирующих моделей. Для наглядности в таблице 4.2 приведены основные частные случаи представления (4.8).

Таблица 4.2 – Основные критерии селекции моделей (частные случаи (4.8))

Критерий	$V(s)$	$\eta_1(W)$	$\eta_2(W)$	σ^2
FPE	RSS	$(n+s)/(n-s)$	0	–
PSE	RSS	1	2s	–
Cp	RSS	1	2s	RSS/(n-m)
AR	$\Delta(B A)$	1	0	–

4.3 Проектирование функциональной и диалоговой структуры программной системы моделирования

Описываемый подход к построению автоматизированной системы структурной идентификации основан на концепции, исходящей из следующих принципов: 1) основные блоки системы соответствуют основным этапам решения задач моделирования по экспериментальным данным (имитация действий естественного интеллекта); 2) различные методы имеют много аналогичных элементов и могут быть обобщены (анализ методов и синтез обобщенного метода на модульном принципе); 3) диалоговая оболочка должна содержать не только формализованные (обоснованные) знания, но и неформальный опыт применения методов (структуризация знаний и опыта); 4) диалог должен иметь несколько уровней сложности с учетом возможного уровня квалификации пользователя (адаптация к пользователю); 5) система должна иметь один или несколько автоматических режимов решения задачи (задача должна быть решена до конца минимум один раз).

Система, базирующаяся на сформулированных принципах, занимает промежуточное положение между диалоговыми ППП и экспертными системами и может быть названа программной системой с элементами искусственного интеллекта [43,44].

Формирование функциональной архитектуры системы. Система моделирования конструируется в виде программного комплекса с иерархической структурой, содержащей пять уровней: основные блоки; функциональные блоки; вычислительные программы; программные модули; программные процедуры.

Верхний уровень системы включает девять основных блоков, соответствующих основным этапам процесса моделирования по экспериментальным данным: старт системы; подготовка данных; подготовка задачи; постановка задачи; формирование задачи; решение задачи; анализ результатов; применение результатов; завершение работы. Их содержательный смысл раскрывается названиями функциональных блоков, обеспечивающих решение следующих групп задач.

Старт системы обеспечивает вхождение в систему и работу с главным меню.

Подготовка данных предполагает формирование и ведение файлов и баз данных и выполнение предварительной обработки данных (сглаживание, заполнение пропусков, масштабирование, центрирование, нормирование и т.д.).

Подготовка задачи состоит в определении типа задачи (моделирование статике, временных рядов или динамики), цели моделирования (аппроксимация, прогнозирование, поиск закономерности).

Для правильной *постановки задачи* необходимы первичный анализ данных (для определения основных статистических характеристик и выявления корреляционных, трендовых и колебательных свойств), определение класса объекта (линейный-нелинейный, статический-динамический, одномерный-многомерный, стационарный-нестационарный и т.д.), выбор класса базисных функций моделей (полиномы, тригонометрические ряды, дискретные разностные уравнения и пр.) и соответствующее преобразование данных.

При *формировании* конкретной задачи моделирования выбираются основной критерий качества моделей, тип генератора структур (переборный, итерационный или неформализованный, назначаемый пользователем), вспомогательный критерий качества оценки параметров (а тем самым метод оценивания), соответствующие программные модули (решатели) и их управляющие параметры.

При *решении* сформированной задачи выполняется собственно процедура моделирования по экспериментальным данным (автоматически, либо при пассивном или активном участии пользователя): в выбранном классе базисных функций по заданному генератору формируется множество моделей-претендентов (с оценкой параметров по вспомогательному (внутреннему) критерию), из которого отбирается одна или несколько моделей, лучших по основному (внешнему) критерию.

Анализ результатов моделирования предполагает верификацию (оценку адекватности) полученных моделей с помощью различных статистических процедур проверки адекватности модели, значимости оценок ее параметров, свойств остатков, вычисления точностных характеристик, смысловой оценки структуры модели.

Под *применением результатов* моделирования может пониматься как интерпретация полученных моделей, оценка их полезности, предоставление их эксперту или лицу, принимающему решения, использование в прогнозных и оптимизационных расчетах или при решении задач управления, так и просто организация хранения полученных результатов, вывод их на печать или визуализация в виде графиков, схем и рисунков.

Решение о завершении работы принимается на основе выводов, полученных в двух предыдущих основных блоках.

Каждый функциональный блок представляет собой относительно самостоятельный пакет вычислительных программ, каждая из которых состоит из головной программы и программных модулей, базирующихся на программных процедурах. Такая иерархическая модульная архитектура программной системы, во-первых, позволяет вести параллельно разработку и программирование блоков, программ, модулей и процедур, а также диалоговых средств, во-вторых, упрощает будущее развитие и совершенствование системы, в-третьих, обеспечивает широкие функциональные возможности системы и ее структурную гибкость при настройке на различные предметные области.

Приведенный проект архитектуры системы моделирования основан на подходе к проблеме структуризации знаний в области моделирования по экспериментальным данным, изложенном в подразделе 4.1. При этом для разработки функциональных блоков системы моделирования существенно используется разработанный выше принцип компоновки различных алгоритмов из стандартного комплекса обобщенных многофункциональных модулей минимальной конфигурации.

Формирование диалогового обеспечения. Современная интерактивная программная система должна содержать, с одной стороны, хорошо структурированную, управляемую и перестраиваемую библиотеку вычислительных модулей, а с другой - целесообразные диалоговые средства, отражающие в систематизированном виде опыт решения конкретных практических задач. Общая структура диалога соответствует основным этапам процесса моделирования и включает необходимое число блоков принятия решений с меню-подобными перечнями вариантов возможных решений, подсказками об их предпочтении и комментариями к их содержанию.

Поскольку система, вообще говоря, должна быть ориентирована на любого пользователя, при разработке концепции диалога в главе 1 принято предположение о том, что *квалификация пользователя* может иметь *три основных уровня*: нулевой, средний и высокий, которым соответствуют названия - начинающий пользователь, подготовленный и эксперт. Тогда на каждом этапе диалога предусматривается три варианта выбора решения: *автоматический, интерактивный* (автоматизированный) и *свободный* (меню-подобный).

Уровень квалификации пользователя можно оценивать с точки зрения навыков работы с ЭВМ, знания конкретной прикладной задачи (объекта) и владения методами моделирования. Если предположить, что система имеет дружественный интерфейс, то требования к навыкам работы с компьютером можно не рассматривать. Тогда минимальный уровень квалификации определяется необходимостью должной подготовки и корректной постановки задачи.

При реализации данной концепции построения диалога, отражающего опыт моделирования, построенная система моделирования может представлять два крайних режима решения задач - чисто автоматический и чисто пакетный (формируемый пользователем-экспертом), а также различные промежуточные варианты режима, естественным образом учитывающие уровень квалификации конкретного пользователя.

Наряду со знаниями, заложенными в диалоговых средствах, интеллектуальный уровень системы определяется также разнообразием сервисных возможностей, зависящих от системных средств ЭВМ и от создания комфортных условий для пользователя с помощью реализации специфических требований, характерных для моделирования.

Общие требования к сервисным возможностям можно разделить на три уровня: минимальные (обязательные), средние (целесообразные) и полные (желательные). К минимальному сервису относятся: автономность (сведения о работе системы содержатся в ней самой); диалог на языке, близком к естественному; разнообразные операции с

данными; визуализация процесса решения; диагностика ошибок пользователя; средства тестирования подсистем; документирование результатов.

Перечень *целесообразных средств* включает: мобильность (переносимость для программно-совместимых ЭВМ); наличие тестовых примеров для характерных типов практических задач; наличие удачно решенных примеров реальных задач с подробным описанием хода решения от данных до результата; интерпретация получаемых результатов с рекомендациями по их улучшению.

К *полным требованиям*, кроме указанных, можно отнести: наличие тест-полигона для имитации типовых вариантов реальных задач с полной характеристикой получаемых результатов; банк типовых примеров (кроме успешно решенных задач, сюда включаются также те, которые не удалось решить вообще); возможность пополнения или замены вычислительных модулей без участия разработчика; стыкуемость с каналами передачи данных, банками данных и другими программными системами; игровой интерфейс и подсистема объяснений или обучения методам моделирования и навыкам их применения.

Изложенный подход к формированию функциональной и диалоговой структуры программной системы моделирования был в той или иной мере использован при разработке программной системы АСТРИД на основе МГУА.

4.4. Общая характеристика возможностей программной системы АСТРИД

Система автоматизированной **структурной идентификации** АСТРИД написана на языке С в системе DOS. Для выполнения системы АСТРИД необходима IBM-совместимая ПЭВМ типа 80286 и выше со стандартной конфигурацией. В существующем варианте системы АСТРИД введены следующие ограничения на объем исходной информации: число аргументов NA не более 50-ти, число точек $N1$ не более 100. Эти ограничения обусловлены быстродействием процессора и объемом памяти. Для ПЭВМ с процессором PENTIUM возможно увеличение предельного числа аргументов и точек.

Прежде, чем приступить к решению задачи моделирования весьма полезно провести визуальный анализ исходной информации: посмотреть графики изменения входных аргументов и выходной величины (программа “Анализ исходных данных”). При этом следует обращать внимание на наличие “диких точек”. Можно также проверить наличие коррелированности аргументов между собой и с выходной величиной (при этом графики следует строить для нормированных значений).

В системе АСТРИД установлены следующие приоритеты автоматического выбора рабочего алгоритма: если $NA < 15$ и $N1 > 2.5 * NA$ - выбирается алгоритм КОМБИ (комбинаторный), если $NA < 51$ и $N1 > 2.5 * NA$ - алгоритм MULTI (многоэтапный), если $NA < 51$ и $N1 > 8$ - алгоритм GMDH (многорядный).

После окончательного выбора модели можно получить расчетные значения выходной величины в дополнительно заданных точках. Значения аргументов могут быть введены с клавиатуры или из подготовленного файла. Можно также посмотреть графики оценок выходной величины и всех или некоторых входных аргументов.

1. Общие сведения о пакете

Пакет программ предназначен для синтеза моделей оптимальной сложности по экспериментальным данным, полученным как в режиме нормальной эксплуатации, так и планируемого эксперимента.

Пакет состоит из программы визуального анализа данных (AID), трех программ моделирования (KOMBI, MULTI и GMDH), программы визуализации результатов моделирования (ARM), программы прогнозирования (PROG) и программы с кратким описанием пакета.

2. Подготовка данных

Файл исходных данных (входной файл) готовится на любом текстовом редакторе в следующем формате:

[NA]	[N1]				
[X1,1]	[X1,2]	...	[X1,h]	...	[X1,N1]
...
[Xi,1]	[Xi,2]	...	[Xi,h]	...	[Xi,N1]
...
[XNA,1]	[XNA,2]	...	[XNA,h]	...	[XNA,N1]
[f1]	[f2]	...	[fh]	...	[fN1]

имяX₁

....

имяX_{NA}

имяf

Разделителями являются пробелы; NA - число аргументов; N1 - число экспериментов; [X_{i,h}], [f_h] - значения i-го аргумента и выходной величины в h-й точке; i = 1,...,NA; h=1,...,N1; имяX_h - название i-о аргумента; имяf - название вых. величины.

Формат ввода : [NA], [N1] - целые; [X_{i,h}], [f_h] - действительные числа; имяX_i, имяf - строка символов.

3. Область применения алгоритмов

NA аргументов представляют исходный базис, из которого формируется рабочий базис (NR аргументов) - входные аргументы, запаздывания от входных аргументов и выходной величины, функции от входных аргументов.

Алгоритм KOMBI позволяет получить линейную относительно рабочего базиса модель, сложность которой (число слагаемых) не превышает NR. Алгоритм может работать при условии, что NR < 15 и N1 > NR.

Алгоритм MULTI позволяет получить линейную относительно рабочего базиса модель, сложность которой (число слагаемых) не превышает NR. Алгоритм может работать при условии, что NR <= 50 и N1 > NR.

Алгоритм GMREZ позволяет получить модель со сложностью выше сложности рабочего базиса. Алгоритм может работать при условии, что $NR < 51$ и $9 < N1 < 100$.

4. Порядок работы с пакетом

1. Подготовить файл исходных данных.
2. Запустить программу PIC1.EXE.
3. В ответ на запросы программы :
 - выбрать файл исходных данных (входной файл) или ввести его имя в виде [patch\\]имя.тип; При этом на экран выводятся названия входных аргументов и выходной величины, таблица их значений.
 - выбрать режим работы : диалоговый или автоматический (в автоматическом режиме алгоритм выбирается программой, в диалоговом - пользователем);
4. По окончании счета автоматически запускается программа ARM.EXE.

На экран будет выведена лучшая из СВ модель, и, после нажатия любой клавиши, графики выходной величины и ее оценок по модели. Пользователь может выбрать удовлетворяющую его модель, которая и является выходом системы.

6. После того, как пользователь выберет удовлетворяющую его модель, можно запустить программу PROG.EXE для вычисления прогнозных значений выходной величины для заданных значений аргументов.

5. Комбинаторный алгоритм (KOMBI)

Комбинаторный алгоритм предназначен для построения полиномиальных моделей оптимальной сложности методом полного перебора. При разделении всей выборки ($N1$ точек) на обучающую (K точек), проверочную ($K1$ точек) и экзаменационную последовательности учитывать ограничения: $K1 \geq 0$, $N1 - K - K1 \geq 0$, $K \geq K1$.

При запуске алгоритм генерирует вектор управляющих параметров:

$$K = 0.6 * N1, K1 = N1 - K, CRI = 0, CB = 3, kl = 0.$$

Умалчиваемые значения могут быть изменены пользователем в диалоговом режиме:

$K, K1$ - длина обучающей и проверочной последовательностей;

$N1$ - число экспериментов;

CRI - ключ критерия: 0 - регулярность; 1 - несмещенность; 2, 3 -

комбинированные; при $K1 = 0$ должно быть $CRI = 0$ - критерием будет ошибка МНК (СКО);

CB - число отбираемых по критерию лучших моделей;

$K1$ - ключ вида полинома (0- полный; 1- без свободного члена; 2- со свободным членом).

6. Переборно-селекционный алгоритм (MULTI)

Переборно-селекционный алгоритм предназначен для построения линейных по параметрам моделей оптимальной сложности с помощью поэтапного усложнения моделей на один аргумент с применением принципа неокончательных решений, при котором усложняется не одна, а несколько (FM) лучших на предыдущем этапе моделей. При разделении всей выборки данных ($N1$ точек) на обучающую (K точек), проверочную ($K1$ точек) и

экзаменационную последовательности требуется учитывать ограничения: $K1 \geq 0$, $N1 - K1 \geq 0$, $K \geq K1$.

При запуске алгоритм генерирует вектор управляющих параметров :

$K = 0.6 * N1$; $K1 = N1 - K$; $CRI = 0$; $CB = 3$; $kl = 0$; $FM = NR$, если $NR \leq 20$; $FM = (20 - \text{entier}[(NR - 20)/2])$, если $20 < NR \leq 50$; $SA = NR$, если $NR \leq 10$, $SA = 10$, если $NR > 10$.

Умалчиваемые значения могут быть изменены пользователем в диалоговом режиме.

$K, K1$ - длина обучающей и проверочной последовательностей;

$N1$ - число экспериментов;

CRI - ключ критерия: 0 - регулярность; 1 - несмещенность; 2,3 - комбинированные; при $K1 = 0$ должно быть $CRI = 0$ - критерием будет ошибка МНК (СКО);

CB - свобода выбора по критерию лучших моделей;

$K1$ - ключ вида полинома (0 - полный; 1 - без свободного члена; 2 - со свободным членом);

SA - наибольшее число этапов алгоритма или наибольшая сложность (число аргументов) сравниваемых моделей;

FM - число лучших моделей, передаваемых на следующий этап алгоритма (свобода выбора на этапе) $SA \leq NR$, $CB \leq FM \leq NR$, $NR \leq 50$.

7. Многорядный алгоритм (GMDH)

Многорядный алгоритм МГУА предназначен для построения линейных ($L=3$) и нелинейных ($L=4,5,6$) моделей оптимальной сложности по экспериментальным данным.

При запуске алгоритм генерирует вектор управляющих параметров: $K = 0.6 * N1$, $K1 = N1 - K$, $CRI = 0$, $L = 6$, $HOP = 0$, $a = 1$, $b = 0$.

Умалчиваемые значения могут быть изменены пользователем в диалоговом режиме.

Для работы алгоритма необходимо выполнение следующих неравенств:

$K + K1 \leq N1$; $K \geq 1.5 * L$; $K1 \geq 3$; $a, b \geq 0$; $a + b > 0$.

$K, K1$ - длина обучающей и проверочной последовательностей;

$N1$ - число экспериментов;

CRI - указатель критерия (0- регулярность (СКО); 1- несмещенность (НС); 2,3- комбинированные);

a, b - весовые коэффициенты комбинированного критерия $KRIT = a * СКО + b * НС$ (обычно используемые значения 0|1);

L - длина частного описания ($L=3,4,5,6$).

8. Программа визуализации моделей (ARM)

Программа визуализации моделей (ARM) позволяет поочередно посмотреть на экране все полученные модели и выбрать лучшую. Любую модель можно вывести на принтер. Для вывода графика на печать нажать клавишу Print Screen (при этом необходимо предварительно загрузить Graphics.com). Для выхода из графического экрана нажать любую клавишу.

9. Программа прогнозирования (PROG)

Программа прогнозирования (PROG) позволяет по отобранной программой ARM модели получить таблицу значений выходной величины для заданных пользователем значений аргументов, которые могут быть введены с клавиатуры по запросу программы (режим К) или считаны из файла исходных данных для прогнозирования (режим F).

По выбору пользователя можно: вывести таблицу на принтер, построить графики прогнозных значений выходной величины и отобранных аргументов.

Файл исходных данных для прогнозирования (*.pro) готовится на любом текстовом редакторе:

[T_n]

[X _{1,1}]	[X _{1,2}]	...	[X _{1,h}]	...	[X _{1,N1}]
...
[X _{i1}]	[X _{i,2}]	...	[X _{i,h}]	...	[X _{i,N1}]
...
[X _{NA,1}]	[X _{NA,2}]	...	[X _{NA,h}]	...	[X _{NA,N1}]

Разделителями являются пробелы; T_h - число точек прогноза; [X_h] - значения i-о аргумента в h-й точке; i = 1,...,NA; h = 1,...,T_n.

Формат ввода : [T_n]- целое; [X_{ih}] - действительные числа.

ВЫВОДЫ

Структурная идентификация рассмотрена как задача поиска модели с минимальной дисперсией ошибки прогнозирования. Разработан метод критических дисперсий, позволивший подробно исследовать закономерности изменения сложности выбираемой структуры в зависимости от уровня шума. Доказан методологически важный результат о необходимости согласовывать сложность модели с уровнем неопределенности в данных.

В данной части отчета получены следующие научные результаты: исследованы закономерности уменьшения сложности (числа оцениваемых параметров) оптимальной модели при увеличении уровня (дисперсии) шума на выходе объекта; разработана методика оценки эффективности применения практических критериев качества структур в задаче минимизации дисперсии ошибки прогнозирования; установлены условия оптимальности или адекватности критериев типа "перекрестного подтверждения", которые используются в МГУА и основываются на разбиении выборки, за счет чего они имеют свойство неявного (автоматического) учета уровня неопределенности в данных.

Разработана методика аналитического оценивания эффективности критериев (до сих пор такая задача решалась на основе вычислительных экспериментов), с помощью которой получены условия адекватного применения критериев МГУА, основанных на разбиении выборки. Предложена постановка задачи, позволившая впервые получить теоретическое обоснование МГУА как метода помехоустойчивого моделирования.

Разработан новый комбинаторно-селекционный алгоритм МГУА с полиномиальной зависимостью объема вычислений от числа переменных вместо экспоненциальной зависимости при полном переборе различных структур. Приведены численные характеристики быстродействия этого алгоритма.

Таким образом, одним из основных научных результатов работы является то, что разработаны основы теории МГУА как метода структурной идентификации помехоустойчивых прогнозирующих моделей. Существенное значение для практики структурной идентификации имеет предложенная методология разработки программного обеспечения, функциональной архитектуры и структуры диалога программных систем моделирования по экспериментальным данным на основе МГУА. Эта методология реализована в диалоговой прикладной программной системе АСТРИД, предназначенной для автоматизированного моделирования и прогнозирования сложных объектов и процессов в разных классах моделей.

Прикладная эффективность МГУА подтверждена решением практических задач моделирования экономических процессов в Украине: бюджетные поступления, объем отраслевой продукции, индекс инфляции, изменение ВВП. Сравнение этих результатов с прогнозирующими свойствами моделей, полученных регрессионным методом, показало преимущество МГУА именно как метода построения прогнозирующих моделей.

ПЕРЕЧЕНЬ ЛИТЕРАТУРЫ

1. Льюнг Л. Идентификация систем. Теория для пользователя. - М.: Наука, 1991.- 432с.
2. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами.- Киев: Техніка, 1975.- 312 с.
3. Ивахненко А.Г., Мюллер И.А. Самоорганизация прогнозирующих моделей. - Киев: Техніка, 1985. - 221с.
4. Себер Дж. Линейный регрессионный анализ.- М.: Мир, 1980. - 456с.
5. Akaike H. Fitting autoregressive model for prediction // Ann. Stat.- 1969.- v.21.- P.243-247.
6. Mallows C.L. Some comments on C_p // Technometrics.- 1973.-v.15.- P.661-667.
7. Акаике Х. Развитие статистических методов / В кн. Современные методы идентификации систем.- М.: Мир, 1983.- С.148-176.
8. Степашко В.С., Кочерга Ю.Л. Методы и критерии решения задач структурной идентификации // Автоматика.- 1985.- N5.-С.29-37.
9. Ивахненко А.Г., Юрачковский Ю.П. Моделирование сложных систем по экспериментальным данным.- М.: Радио и связь, 1986.- 118с.
10. Любич Ю.И., Майстровский Г.Д. Общая теория релаксационных методов для выпуклых функционалов // Успехи мат.наук.- 1970.- N 25.- вып.1.- С.57-112.
11. Степашко В.С., Кочерга Ю.Л. Классификация и анализ помехоустойчивости внешних критериев селекции моделей // Автоматика. - 1984.- N 4.- С.38-50.
12. Степашко В.С. Асимптотические свойства внешних критериев селекции моделей // Там же. - 1988. - N6.- С.75-82.
13. Демиденко Е.З. Линейная и нелинейная регрессии. - М.: Финансы и статистика, 1981. - 302 с.
14. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. - М.: Наука, 1979.- 447с.
15. Ивахненко А.Г., Карпинский А.М. Самоорганизация моделей на ЭВМ в терминах общей теории связи (теории информации) // Автоматика.- 1982.- N 4.- С.7-26.
16. Котельников В.А. Теория потенциальной помехоустойчивости. - М.: Госэнергоиздат, 1956. - 253с.
17. Игнатов В.А. Теория информации и передачи сигналов. - М.: Сов.радио, 1979. - 280с.
18. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. - Киев: Наук.думка, 1985. - 216с.
19. Шеннон К. Работы по теории информации и кибернетике. - М.: Изд-во иностр.лит., 1963. - 828с.
20. Степашко В.С. Структурная идентификация прогнозирующих моделей в условиях планируемого эксперимента // Автоматика.- 1992. - N 1. - С.26-35.
21. Степашко В.С. Селективные свойства критерия непротиворечивости моделей // Там же.- 1986.- N 2.- С. 40-49.
22. Павлов О.А. Критерій ранжування для порогового самовідбору змінних в алгоритмах МГВА // Там же.- 1969.- N 4.- С. 89-91.

23. Ивахненко О.Г., Коппа Ю.В., Тодуа М.М., Петраке Г. Метод математичного моделювання складних екологічних систем // Там же.- 1971.- N 4.- С.20-34.
24. Шелудько О.І. Алгоритм МГВА з ортогоналізованим частинним описом для синтезу моделей за результатами планованого експерименту.- Там же.- 1974.- N5.- С.32-42.
25. Висоцький В.М. Про найкращий поділ вихідних даних в алгоритмах МГВА // Там же.- 1976.- N 3.- С.71-74.
26. Юрачковский Ю.П., Грошков А.Н. Оптимальное разбиение исходных данных на обучающую и проверочную последовательности на основе анализа функции распределения критерия // Там же.- 1980.- N 2.- С. 5-9.
27. Сарычев А.П. Решение проблемы разбиения в МГУА при расчете критерия регулярности в условиях активного эксперимента // Там же. - 1989. - N 4. - С.19-26.
28. Раскин В.Г., Яновский Л.П. О выборе регрессионной модели оптимальной сложности с помощью критерия регулярности // Там же.- 1984.- N 1.- С.40-43.
29. Кочерга Ю.Л. Теоретический критерий эффективности прогноза // Там же.- 1986.- N 4.- С.18-123.
30. Справочник по типовым программам моделирования. Под ред. А.Г.Ивахненко, – К.: Техніка, 1980.-184 с.
31. Бюлетень економічної кон'юнктури України. – Київ: НДІ статистики Міністату України. – 1997. - випуск N 3. - 134 с.
32. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия: В 2-х вып. Вып. 2. – М.: Финансы и статистика, 1982. – 239с.
33. Тенденції української економіки. Місячний бюлетень. Європейський центр макроекономічного аналізу.- Квітень 1995.
34. Павловський М.А. Шлях України. - Київ: Техніка, 1996.- 152 с.
35. Костіна Н.І. Алексеев А.А. Василик О.Д. Фінансове прогнозування: методи та моделі.- Київ: Знання, 1997.
36. Ивахненко А.Г., Ивахненко Г.А. Нормативный прогноз и оптимальное управление многомерными объектами при помощи самоорганизации системы нефизических моделей // Проблемы управления и информатики.- 1996.- № 1-2.
37. Madala H.R., Ivakhnenko A.G. Inductive Learning Algorithms for Complex Systems Modeling.- New York: CRC Press Inc., Boca Raton.- 1994.- 384 p..
38. Белогуров В.П. Критерий пригодности моделей для прогноза процессов // Автоматика.- 1990.- N 3.- С.23-28.
39. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. – М.: Финансы и статистика, 1987. – 239с.
40. Енюков И.С. Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА.-М.: Финансы и статистика, 1986. - 232с.
41. Сильвестров Д.С. Программное обеспечение прикладной статистики.- М.: Финансы и статистика, 1988.- 240 с.

42. Степашко В.С. Алгоритмы МГУА как основа автоматизации процесса моделирования по экспериментальным данным // Автоматика.- 1988.- N 4.- С.44-55.
43. Элти Дж., Кумбс М. Экспертные системы: концепции и примеры / Пер. с англ. и предисл. Б.И.Шитикова. - М.: Финансы и статистика, 1987. - 191 с.
44. Hahn G.I. More intelligent statistical software and statistical expert systems: future directions // Amer.Stat.- 1985.- v.39.- N 1.- P. 1-16.