
МЕТОДИЧЕСКИЕ АСПЕКТЫ ПОСТРОЕНИЯ ДИСКРЕТНО- НЕПРЕРЫВНЫХ МОДЕЛЕЙ ПОКАЗАТЕЛЕЙ ЭКОЛОГО- ЭКОНОМИЧЕСКИХ СИСТЕМ

Давнис Валерий Владимирович,

доктор экономических наук, заведующий кафедрой информационных технологий и математических методов в экономике Воронежского государственного университета; vdavnis@mail.ru

Умывакин Василий Митрофанович,

доктор географических наук, профессор кафедры математических методов исследования операций Воронежского государственного университета; umyvakin@mail.ru

Климов Алексей Валерьевич,

аспирант Воронежского института экономики и права;
aleksey@alfey.net

Предлагается подход к построению статистических зависимостей критериев качества и управляющих параметров эколого-экономических систем на основе оригинального метода дискретно-непрерывного моделирования показателей.

Ключевые слова: эколого-экономическая система, анализ многомерных данных, дискретно-непрерывные модели показателей

В прикладных исследованиях эколого-экономических систем (ЭЭС) активно применяются методы анализа многомерных данных (методы регрессионного анализа, кластер-анализа, многомерного шкалирования и др.) [1 – 3]. Однако используемый при этом стандартный аппарат прикладной статистики дает адекватное описание изучаемых явлений при выполнении целого ряда довольно жестких априорных требований. На практике некоторые из этих предположений оказываются невыполнимыми (в частности, гипотеза о нормальности распределения), что существенно ограничивает сферу применения моделей анализа данных (в первую очередь, моделей регрессионного анализа). Методы кластер-анализа (многомерной классификации) позволяют выявлять только структурные закономерности и не дают достаточно точного формализованного описания исследуемых ЭЭС.

Поэтому наибольшего эффекта в экономико-статистическом моделировании можно достигнуть при совместном применении методов классификации и корреляционно-регрессионного анализа. Следует отметить, что основная роль методов кластер-анализа заключается в том, что их применение позволяет получить однородные классы ЭЭС (объектов) и тем самым обеспечить построение в каждом классе адекватных статистических моделей с помощью методов регрессионного анализа. Такое совместное применение методов называют методом дискретно-непрерывного моделирования показателей [4], а систему регрессионных моделей для всех классов – дискретно-непрерывной моделью показателей ЭЭС.

Дискретно-непрерывная модель (ДНМ), как следует из самого определения, представляет собой комбинацию моделей двух типов – дискретной и непрерывной. При ее использовании совокупность анализируемых объектов моделируется на двух уровнях – уровне классов и уровне отдельных объектов. Модели верхнего уровня (уровень классов) дискретные; в общем случае они строятся методами классификации. Модели нижнего уровня – непрерывные и представляют собой внутриклассовые уравнения регрессии. Таким образом, исследуемая статистическая зависимость может быть представлена не одной моделью, а набором регрессионных уравнений, каждое из которых действительно в определенной, строго ограниченной области пространства показателей ЭЭС.

Рассмотрим два подхода к построению дискретно-непрерывных моделей (ДНМ) показателей состояния сложных систем, каждый из которых предполагает наличие качественных факторов, включение которых в непрерывную модель по тем или иным причинам затруднительно.

Первый подход предложил Б.Б. Розин [4]. В работе [5] он называется методом типологической регрессии. Данный подход, заключается в первоначальном выделении (например, с помощью методов типологической группировки [4] или кластер-анализа [1,3]) по некоторым нечисловым и/или количественным показателям групп (типов, классов) объектов, для которых затем строятся регрессионные уравнения [2]. Они должны существенно отличаться друг от друга. При этом значения моделируемого показателя, в основном, определяются различиями в значениях параметров в разных группах.

В работе [4] задачу построения ДНМ показателей сложных систем предлагается решать в рамках следующей схемы:

- 1) обоснование гипотезы о неоднородности совокупности и целесообразности использования дискретно-непрерывной модели;
- 2) построение дискретной части модели, включающее выбор способа классификации и набора показателей, по которым должно производиться разбиение совокупности на однородные классы, формирование математического описания классов и правила отнесения нового объекта к

тому или иному классу, оценку точности классификации;

3) построение непрерывной части модели, включающее все операции, присущие регрессионному анализу, но выполняемые для каждого класса в отдельности;

4) оценка аппроксимирующей способности модели.

Здесь особое значение имеют результаты проверки гипотезы о неоднородности совокупности анализируемых объектов и исследования вопроса соизмеримости показателей различных типов. Один из возможных способов проверки гипотезы о неоднородности совокупности заключается в использовании многомерной классификации с последующей оценкой существенности различия полученных в результате разбиения классов. Дело в том, что большинство алгоритмов кластер-анализа базируется на выделении компактных классов с приближенно-нормальным многомерным распределением показателей. Таким образом, если в результате разбиения получены существенно различные классы, то есть основание принять гипотезу о том, что исходная совокупность неоднородна и представляет собой смесь, композицию из нескольких более однородных подмножеств. Существенно различными классы считаются в том случае, если существенно различаются основные статистические характеристики каждого из показателей в отдельности (параметры кривых распределения) и характеристики их взаимосвязи (матрицы коэффициентов корреляции): Такой анализ может проводиться либо путем попарного сравнения классов между собой, либо путем сравнения каждого класса с общей совокупностью.

При построении дискретной части ДНМ прежде всего определяется, надо ли проводить разбиение только в пространстве параметров (факторов-аргументов) или и в объединенном пространстве моделируемого критерия качества (фактор-функции) и параметров ЭЭС. Если ставится цель построения ДНМ с линейными уравнениями регрессии (а это наиболее распространенная постановка задачи), то это равносильно требованию формирования классов с многомерным распределением показателей, близким к нормальному.

После того как набор показателей для классификации объектов (ЭЭС) сформирован, можно переходить к выбору способа разбиения совокупности объектов на группы. Следует различать два принципиально различных способа классификации: метод последовательных логических делений по каждому показателю в отдельности (типологическую группировку) и метод формирования групп на основе набора показателей (многомерная классификация), ни один из которых не является необходимым или достаточным условием принадлежности объекта к данной группе.

При использовании любого из методов многомерной классификации один из важных вопросов построения дискретной части ДНМ – определение рационального числа классов. С точки зрения способа определения числа классов все алгоритмы кластер-анализа можно подразделить на две

группы: алгоритмы, где число классов заранее задается исследователем, и алгоритмы, где число классов в явной форме не задается, а находится в процессе разбиения. В этом подходе при построении ДНМ очень важна задача отыскания небольшого набора показателей, которые объясняли бы основную часть внутриклассовой вариации моделируемого критерия. Начальным этапом этого процесса является анализ структуры матриц парных коэффициентов корреляции и выделение взаимосвязанных групп показателей. Далее исключаются дублирующие и тесно коррелированные показатели, производится выбор существенных показателей для построения частных регрессий. Наиболее распространенные формальные способы отыскания существенных показателей базируются на последовательном исключении показателей, дающих наименьшее (наибольшее) изменение остаточной дисперсии по сравнению с исходным набором (например, с помощью метода главных компонент [3]).

Отметим, что в методе типологической регрессии [5] второй и третий этапы в приведенной выше схеме построения ДНМ находятся в соотношении итерационного взаимодействия, то есть непрерывно из очередной итерации второго этапа переходят в следующую итерацию третьего и обратно до тех пор, пока не определятся классификация и коэффициенты регрессии, удовлетворяющие некоторому критерию качества решения данной задачи.

В основе предлагаемого подхода [6 – 8] к построению ДНМ показателей ЭЭС лежит совместное применение двух методов статистического анализа данных – метода регрессионного анализа и метода главных компонент [1,3]. Метод главных компонент позволяет проводить классификацию анализируемых объектов в пространстве новых некоррелированных переменных и является распространенным методом линейного многомерного шкалирования (визуализация данных) [3].

Пусть из всех показателей (переменных), характеризующих ЭЭС, выделен набор управляемых переменных – критериев качества (факторов-функций) $y=(y_1, y_2, \dots, y_M)$ и набор управляющих переменных – параметров (факторов-аргументов) $x=(x_1, x_2, \dots, x_L)$. Обозначим через y_{j_0} анализируемый (моделируемый) критерий качества, статистически связанный с параметрами, $1 \leq j_0 \leq M$. Предположим также, что имеется некоторая совокупность (выборка) из N объектов, для которых известны значения величин $y_{j_0}^i$ и x_k^i i -го объекта, $k=1, 2, \dots, L$; $i=1, 2, \dots, N$. Требуется установить регрессионную зависимость y_{j_0} от x_1, x_2, \dots, x_L .

Разделение совокупности анализируемых объектов (ЭЭС) на четко обособленные группы и построение регрессионных моделей для каждого из них во многих практических ситуациях является затруднительным. Например, когда невозможно разбиение объектов на группы по непересекающимся интервалам диапазона изменения параметров (рис. 1).

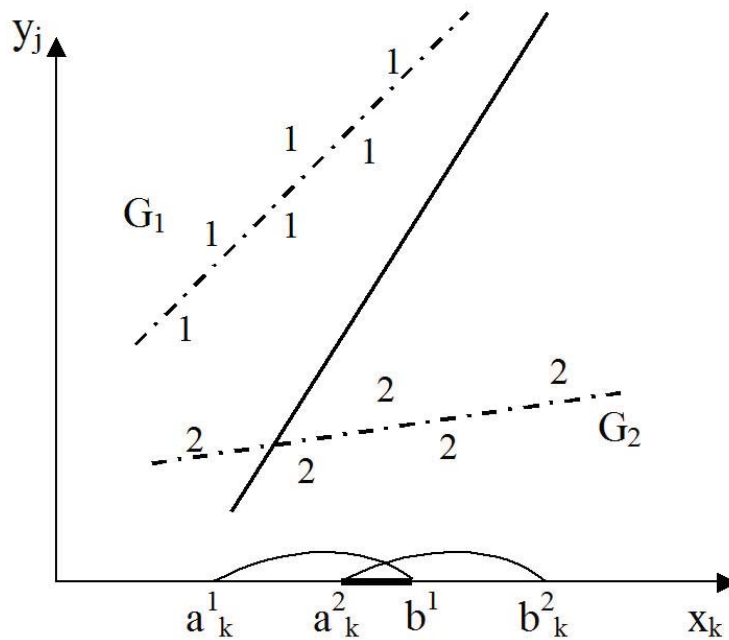


Рис.1. Классификация объектов по регрессионным моделям анализируемого критерия качества

- 1 и 2 – объекты (точки) соответственно классов G_1 и G_2 ;
 $[a_k^1, b_k^1]$ и $[a_k^2, b_k^2]$ – диапазоны изменения параметра x_k в классах G_1 и G_2 ;
 |————| – общий интервал изменения параметра x_k для классов G_1 и G_2 ;
 ————— – уравнение регрессии для всей совокупности объектов;
 - - - - - – уравнения регрессии для классов G_1 и G_2 .

В данном случае применим другой подход [6 – 8]. Он основан на предположении, что исходная информация, необходимая для построения ДНМ содержится в самом моделируемом критерии качества ЭЭС. Так, для объектов с близкими значениями параметров значения этого критерия могут сильно отличаться. Это позволяет сделать вывод, что измерения показателей сделаны на объектах из разных групп, которым соответствуют существенно различные уравнения регрессии – «центры» групп (рис. 1.). Здесь можно рассмотреть вопрос о качестве такой группировки, который решается с помощью проверки статистической гипотезы об идентичности уравнений регрессии [2]. Если гипотеза отвергается, то есть основания считать, что разбиение совокупности объектов на группы проведено верно.

Для предлагаемого подхода к построению ДНМ, который представляет собой некоторый гибрид методов классификации и регрессии, характерна ситуация, когда достоверность числовых параметров, описывающих качественные факторы, недостаточна или их измерение практически не осуществимо. Естественно, получение объективной классификации в этом случае невозможно.

Как известно, важнейшими характеристиками качества (достоверности)

уравнений регрессии являются выборочный коэффициент множественной корреляции R и дисперсионное отношение Фишера F . Рассматриваемый метод построения ДНМ показателей ЭЭС, по существу, направлен на повышение качества уравнений регрессии в каждой группе. Следовательно, исследуемая статистическая зависимость критериев качества и параметров ЭЭС может быть представлена не одной непрерывной моделью, а набором уравнений регрессии, каждое из которых применимо только внутри определенной группы объектов, соответствующей строго ограниченной области пространства параметров.

С формальной точки зрения требуется разбить всю совокупность (множество) G анализируемых объектов на две группы G_1 и G_2 , в каждой из которых коэффициенты множественной корреляции R_1 и R_2 выше, чем коэффициент R для всей выборки. Приведем общее описание алгоритма метода. Вначале строится линейная регрессионная модель анализируемого критерия и определяются выборочные значения R и F для всего множества объектов. Если полученная модель является адекватной и статистические данные таковы, что $R \geq R^*$, где R^* – требуемое значение коэффициента множественной корреляции, разбиение множества G на группы G_1 и G_2 не производится. Предположим, что $R < R^*$, т.е. требования к качеству уравнений регрессии не выполняются. В данном случае метод построения ДНМ показателей ЭЭС позволяет разделить всю совокупность объектов на две группы G_1 и G_2 , в которых соответствующие значения коэффициентов множественной корреляции R_1 и R_2 увеличиваются по сравнению с R . В группах G_1 и G_2 строятся соответствующие уравнения регрессии и определяются коэффициенты множественной корреляции R_1 и R_2 , которые сравниваются с требуемым значением R^* . Если ограничение $R_m \geq R^*$ ($m=1,2$) не выполняется, группы первого иерархического уровня (G_1 и G_2) разбиваются на группы второго уровня и т.д. (рис. 2.).

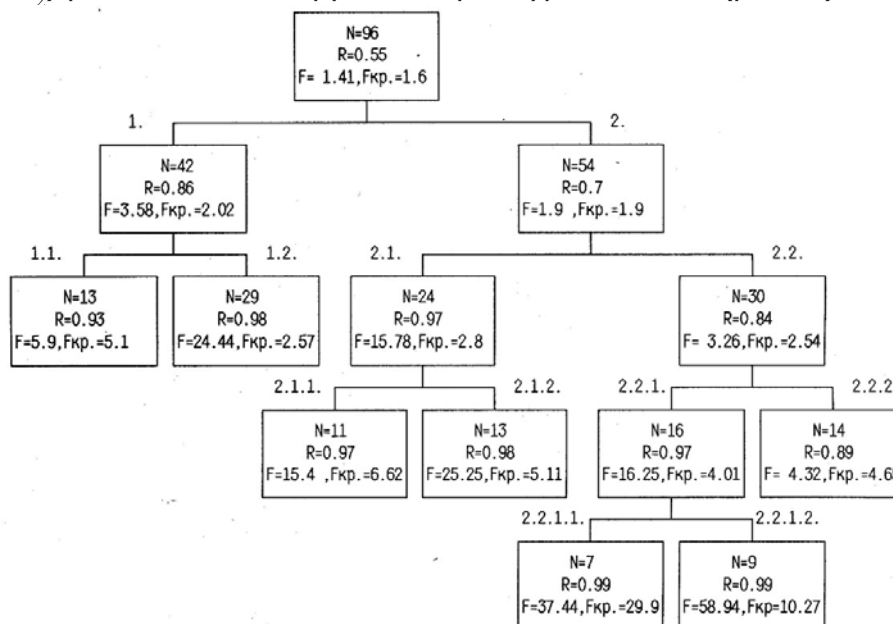


Рис. 2. Построение дискретно-непрерывной модели критерия качества эколого-экономических систем

Рассмотренный метод построения ДНМ применялся для статистического моделирования взаимосвязей между критериями и параметрами ЭЭС при обосновании инвестиций в комплекс программных природоохранных мероприятий по борьбе с водной эрозией земель в ЦЧР. На рис. 2. приведено иерархическая структура процесса разбиения на группы речных водосборов (ЭЭС) по ДНМ моделируемого критерия y_{j0} – «смытость почв с площади земель сельскохозяйственного назначения, %».

Для всей совокупности объектов (96 речных водосборов ЦЧР) линейная регрессионная зависимость между критерием y_{j0} и параметрами x_1 («облесенность территории, %») и x_2 («глубина местного базиса эрозии, м») недостаточно достоверна ($F=1.41$, $R=0.55$). Использование программной системы, реализующей метод построения ДНМ показателей ЭЭС, позволило выделить две группы объектов первого иерархического уровня разбиения (содержащих соответственно 42 и 54 объекта), четыре группы второго уровня, четыре третьего и две четвертого (рис. 2.). В каждой группе было построено свое уравнение регрессии (табл.).

Таблица

Дискретно-непрерывные модели критерия «смытость почв»
для речных водосборов ЦЧР

Группы водосборов	Средние значения	Корреляционная матрица	Уравнения регрессии в стандартизованном и натуральном масштабе
1.1.	$\bar{y}_1 = 30.82$ $\bar{x}_1 = 13.16$ $\bar{x}_2 = 97.15$	1.00 -0.48 0.85 -0.48 1.00 -0.14 0.85 -0.14 1.00	$z_1 = -0.36t_1 + 0.80t_2$ $y_1 = -6.04 - 0.77x_1 + 0.48x_2$
1.2.	$\bar{y}_1 = 23.11$ $\bar{x}_1 = 11.55$ $\bar{x}_2 = 112.34$	1.00 -0.35 0.94 -0.35 1.00 -0.09 0.94 -0.09 1.00	$z_1 = -0.27t_1 + 0.91t_2$ $y_1 = -16.24 - 0.64x_1 + 0.41x_2$
2.1.1.	$\bar{y}_1 = 21.85$ $\bar{x}_1 = 5.93$ $\bar{x}_2 = 120.99$	1.00 0.75 0.97 0.75 1.00 0.71 0.97 0.71 1.00	$z_1 = 0.12t_1 + 0.88t_2$ $y_1 = -16.58 + 0.51x_1 + 0.29x_2$
2.1.2.	$\bar{y}_1 = 24.24$ $\bar{x}_1 = 10.89$ $\bar{x}_2 = 137.77$	1.00 0.56 0.98 0.56 1.00 0.58 0.98 0.58 1.00	$z_1 = -0.02t_1 + 0.99t_2$ $y_1 = -18.5 - 0.11x_1 + 0.31x_2$
2.2.1.	$\bar{y}_1 = 14.87$ $\bar{x}_1 = 11.47$ $\bar{x}_2 = 133.06$	1.00 -0.58 0.93 -0.58 1.00 -0.35 0.93 -0.35 1.00	$z_1 = -0.29t_1 + 0.84t_2$ $y_1 = -7.58 - 0.38x_1 + 0.20x_2$
2.2.2.	$\bar{y}_1 = 12.64$ $\bar{x}_1 = 6.84$ $\bar{x}_2 = 143.64$	1.00 0.06 0.89 0.06 1.00 0.19 0.89 0.19 1.00	$z_1 = -0.11t_1 + 0.91t_2$ $y_1 = -13.37 - 0.20x_1 + 0.19x_2$

Коэффициенты множественной корреляции увеличились в группах до 0.89-0.99, а отношение Фишера до 4.32-58.94. Все полученные регрессионные модели являются адекватными при уровне значимости $\alpha=0.01$. Исключение составляет лишь группа 2.2.2. Для объектов данной группы можно использовать уравнение регрессии группы 2.2.

Список источников

1. Айвазян, С.А. Классификация многомерных наблюдений [текст] / С.А. Айвазян, З.И. Бежаева, О.В. Староверов. – М.: Статистика, 1974 – 154 с.
2. Айвазян, С.А. Прикладная статистика: Исследование зависимостей [текст]: Справ. изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1989. – 487 с.
3. Прикладная статистика: Классификация и снижение размерности [текст]: Справ. изд. / С.А. Айвазян и [др.]. – М.: Финансы и статистика, 1989. – 607 с.
4. Розин, Б.Б. Теория распознавания образов в экономических исследованиях [текст] / Б.Б. Розин. – М.: Статистика, 1973. – 224 с.
5. Азарян, О.В. Алгоритмы типологической регрессии в задачах автоматизации процесса проектирования (на примере создания и комбинирования проектных решений в гидростроительстве) [текст] / О.В. Азарян // Статистика. Вероятность. Экономика: Ученые записки по статистике. – М.: Наука, 1985. – Т.49 – С. 69 – 77.
6. Давнис, В.В. Эвристический метод построения дискретно-непрерывных моделей экономических показателей / В.В. Давнис, И.Б. Русман, Б.Е. Штейн [текст] // Вопросы оптимального программирования в производственных задачах. – Воронеж, 1974. – С. 90 – 94.
7. Умывакин, В.М. Интегральная эколого-хозяйственная оценка и управление земельными ресурсами в регионе [текст] / В.М. Умывакин – Воронеж: ВГПУ, 2002. – 178 с.

METHODICAL ASPECTS OF CONSTRUCTION OF DISCRETE AND CONTINUOUS MODELS OF INDEXES OF ECOLOGICAL AND ECONOMICAL SYSTEMS

Davnis Valeriy Vladimirovich,

Dr. Sc. of Economy, Chief of the Chair of Information Technologies and
Mathematical Methods of Voronezh State University;

Umyvakin Vasilii Mitrofanovich,

Dr. Sc. of Geography, Professor of the Chair of Information Technolo-
gies and Mathematical Methods of Voronezh State University;
umyvakin@mail.ru

Klimov Aleksey Valeryevich,

Post-graduate student of Voronezh Institute of Economy and Law;
aleksey@alfey.net

Approach to construction of statistical dependency of quality criteria
and control parameters of ecological and economic systems based on
the original method of discrete-continuous simulation performance is
offered.

Keywords: ecological an economical system, analyses of multiple
data, discrete and continuous models of indexes.