УДК 519.95:330.115

ФАКТОРНОЕ МОДЕЛИРОВАНИЕ ДЕЯТЕЛЬНОСТИ ФУНКЦИОНАЛЬНЫХ ПОДРАЗДЕЛЕНИЙ ПРОИЗВОДСТВЕННО-ФИНАНСОВЫХ СТРУКТУР

к.т.н. В.Г. Кучмиев, к.т.н. А.И. Лысенко
(представил д.т.н., проф. В.М. Вартанян)

Рассматриваются проблемы формализации на основе статистических данных зависимости результирующего показателя функциональной деятельности от обуславливающих этот результат технико-экономических факторов.

Любое функциональное подразделение производственно - экономической или финансово - хозяйственной направленности характеризуется некоторой структурой, определенной технологией деятельности, заданным количеством исходных и вновь вводимых ресурсов.

Формально все эти параметры, определяющие функционирование структурного подразделения, могут быть отражены в факторной модели типа производственной функции, которая создается на основе статистических данных с помощью методов регрессионного и корреляционного анализов.

Отражая причинно следственную зависимость результирующего показателя от обуславливающих его характеристик, факторная модель позволяет установить: как в среднем изменяется случайная величина результирующего показателя с изменением одной или нескольких случайных величин, обуславливающих факторов в условиях, когда действует большое количество таких факторов и ряд из них неизвестен или просто не учитывается.

Построение факторной модели $\mathbf{k_0 = f(k_1, ..., k_m)}$, отражающей зависимость результирующего показателя $\mathbf{k_0}$ от характеризующих его факторов $\mathbf{(k_1, ..., k_m)}$, состоит из следующих этапов:

1) спецификация модели;
2) формирование статистических данных;
3) параметризация модели.

Этап спецификации модели включает в себя следующие задачи:

1) выбор основных факторов $\mathbf{(k_1, ..., k_m)}$, характеризующих заданный показатель $\mathbf{k_0}$;

2) выбор вида формальной зависимости заданного показателя $\mathbf{k_0}$ от выбранных факторов $\mathbf{(k_1, ..., k_m)}$.

Выбираемые факторы $\mathbf{(k_1, ..., k_m)}$ должны отвечать следующим требованиям:

1) факторы должны быть количественно измеримы, т.е. иметь единицу измерения;

2) факторы должны быть независимы между собой, т.е. в модель не должны одновременно включаться либо факторы, находящиеся между собой в строгой функциональной зависимости, либо факторы, между которыми существует тесная корреляционная связь;

3) факторы должны рассчитываться по данным существующей статистики (отчётности) и этих данных должно быть достаточное количество; считается, что число наблюдений должно, по меньшей мере, в 5-6 раз превышать количество неизвестных параметров модели.

При выборе конкретного вида формальной зависимости связи

$$\mathbf{k_0 = f(k_1, ..., k_m)}$$

необходимо, во-первых, учитывать адекватность математического отображения существующей зависимости и, во-вторых, принимать во внимание удобство оперирования с выбранной формальной функцией с точки зрения целей моделирования.

Как правило, в технико-экономических факторных моделях наибольшее распространение благодаря своей простоте и удобству оперирования получила следующая степенная зависимость:

$$\mathbf{k_0 = A\prod_{i=1}^{m} k_i^{a_i}},$$

где $\mathbf{A}$ - параметр (коэффициент пропорциональности), отражающий влияние на результирующий показатель $\mathbf{k_0}$ неучтённых в модели факторов; $\mathbf{a_i, i = \overline{1,m}}$ – параметры модели (коэффициенты регрессии), отражающие степень влияния каждого выбранного фактора $\mathbf{k_i, i \in \{1,m\}}$ на результирующий показатель $\mathbf{k_0}$.

При формировании статистических данных используют следующие виды наблюдений.

1. Единовременные или перекрестные наблюдения.

В этом случае статистические данные выбираются за один отчётный период времени по различным, но качественно однородным в отношении выбранных признаков, объектам наблюдения. При этом количество наблюдений соответствует количеству обследуемых объектов.

В этом случае факторная модель получается универсальной, т.е. пригодной для описания всех обследуемых объектов из рассматриваемой совокупности, но проигрывает в точности отражения специфики каждого отдельно взятого объекта.

2. Разновременные наблюдения. Представляют собой статистические данные, взятые за различные периоды времени по одному конкретному объекту исследования из рассматриваемой совокупности. В этом случае количество наблюдений соответствует количеству рассмотренных отчётных пе-

риодов по данному объекту. Полученные статистические данные, так называемые временные ряды, хорошо отражают специфику наблюдаемого объекта. Факторная модель получается более специализированной, т.е. более точно отражающей специфику моделируемого объекта, но проигрывает в универсальности, так как не пригодна для моделирования других объектов из рассматриваемой качественно однородной совокупности.

3. Смешанные наблюдения. В этом случае статистические данные составляют из комбинации перекрестных и разновременных наблюдений. Общее количество наблюдений соответствует матрице по объектам за различные отчётные периоды. Факторная модель, полученная на основе смешанных наблюдений, представляет собой компромиссный вариант между специализированной и универсальной моделями.

Результатом формирования статистических данных является полученная статистическая совокупность наблюдений.

Этап параметризации модели заключается в расчёте численных значений параметров $\mathbf{A}, \mathbf{a_i}, \mathbf{i} = \overline{\mathbf{1,m}}$ выбранной формальной зависимости

$$\mathbf{k_0 = A\prod_{i=1}^{m} k_i^{a_i}}$$

на основе исходных статистических данных с помощью так называемых нормальных уравнений.

Наиболее распространённым приёмом получения системы нормальных уравнений является метод наименьших квадратов, который применим для зависимостей

$$\mathbf{z = a_0 + \sum_{i=1}^{m} a_i u_i},$$

линейных относительно своих неизвестных параметров $\mathbf{(a_0, a_1, ..., a_n)}$. Здесь $\mathbf{u_i}, \mathbf{i} = \overline{\mathbf{1,m}}$ - любые функции одного или нескольких аргументов - факторов $\mathbf{k_i}, \mathbf{i} \in \{\mathbf{1,m}\}$, не содержащие неизвестных параметров $\mathbf{(a_i)}, \mathbf{i = 0, 1, ..., m}$.

Для предлагаемой зависимости

$$\mathbf{k_0 = A\prod_{i=1}^{m} k_i^{a_i}}$$

линейная форма будет иметь вид

$$\mathbf{\mathit{l}n\, k_0 = \mathit{l}n\, A + \sum_{i=1}^{m} a_i\, \mathit{l}n\, k_i}.$$

Используя метод наименьших квадратов, получим следующую линейную систему $\mathbf{(m+1)}$ уравнений относительно $\mathbf{(m+1)}$ неизвестных параметров $\mathbf{(a_0, a_1, ..., a_m)}$:

$$\sum_{j=1}^{n} z_j = na_0 + a_1\sum_{j=1}^{n} u_{1_j} + a_m\sum_{j=1}^{n} u_{m_j};$$

$$\sum_{j=1}^{n} z_j u_{1_j} = a_0\sum_{j=1}^{n} u_{1_j} + a_1\sum_{j=1}^{n} u_{1_j}^2 + \ldots + a_m\sum_{j=1}^{n} u_{m_j} u_{1_j};$$

$$\ldots \qquad \ldots \qquad \ldots \qquad \ldots$$

$$\sum_{j=1}^{n} z_j u_{m_j} = a_0\sum_{j=1}^{n} u_{m_j} + a_1\sum_{j=1}^{n} u_{1_j} u_{m_j} + \ldots + a_m\sum_{j=1}^{n} u_{m_m}^2,$$

где $z_j = ln\, k_{0j}, j = \overline{1,n};$ $u_{ij} = ln\, k_{ij}, i = \overline{1,m};$ $a_0 = ln\, A;$ n – число наблюдений в выбранной совокупности.

Решение полученной системы линейных уравнений определяет значения искомых параметров факторной модели $k_0 = A\prod_{i=1}^{m} k_i^{a_i}$, где $A = e^{a_0}$.

Для оценки корреляционной существенности выбранного вектора аргументов - факторов, характеризующих результатирующий показатель полученной модели, рассчитывается матрица парных коэффициентов корреляции:

$$\begin{pmatrix} r_{zz}, & r_{zu_1}, & \ldots, r_{zu_m} \\ r_{u_1 z}, & r_{u_1 u_1}, & \ldots, r_{u_1 u_m} \\ \vdots & \vdots & \vdots \\ r_{u_m z}, & r_{u_m u_1}, & \ldots, r_{u_m u_m} \end{pmatrix},$$

где $z_{u_i u_j} = \dfrac{\sum_{k=1}^{n}\left(u_{ik} - \overline{u_i}\right)\left(u_{jk} - \overline{u_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(u_{ik} - \overline{u_i}\right)^2 \sum_{k=1}^{n}\left(u_{jk} - \overline{u_j}\right)^2}};$ $\quad \overline{u_i} = \dfrac{\sum_{k=1}^{n} u_{ik}}{n};$ $\quad \overline{u_j} = \dfrac{\sum_{k=1}^{n} u_{jk}}{n}$

и определяется коэффициент множественной корреляции

$$R = \sqrt{1 - \Delta/\Delta_{zz}},$$

где Δ – определитель матрицы парных коэффициентов корреляции, а Δ_{zz} – алгебраическое дополнение его элемента r_{zz}.

Парные коэффициенты корреляции служат для определения наиболее значимых в корреляционном отношении аргументов - факторов по отношению к результирующему показателю согласно условию

$$\left|r_{zu_i}\right| \geq 0{,}8.$$

Независимость аргументов-факторов между собой характеризуется условием

$$\left|r_{u_i u_j}\right| < 0{,}8.$$

Коэффициент множественной корреляции, изменяясь в пределах $0 \le R \le 1$ характеризует полноту охвата существенных аргументов-факторов, включённых в модель. Близость коэффициентов множественной корреляции к единице свидетельствует о том, что в полученную модель включены все наиболее существенные факторы, обуславливающие результирующий показатель.

Получив факторную модель в виде конкретной формальной зависимости с вычисленными параметрами, необходимо проверить её адекватность по отношению к реально существующей зависимости с целью установления степени точности и надежности получаемых с её помощью оценок.

Для этого проводится анализ так называемых статистических ошибок, которые неизбежно возникают из-за самой сущности корреляционной зависимости и применяемых методов её нахождения.

Характерные для факторных моделей статистические ошибки можно с известной долей условности разделить на три следующие группы:

1) ошибки выбранной формальной зависимости;

2) ошибки выборки;

3) ошибки наблюдения.

Ошибки полученной формальной зависимости заключаются в том, что для каждого наблюдаемого сочетания выбранных аргументов - факторов расчётная величина результирующего показателя не совпадает точно с его фактическим значением за исключением редких и случайных совпадений.

Величина отклонения $\left(z_j - z_j^p\right)$, представляющая собой ошибку уравнения связи

$$z_j^p = a_0 + \sum_{i=1}^{m} a_i u_{ij}; \;\; \forall j \in \left\{\overline{1,n}\right\}$$

характеризуется величиной средней квадратической ошибки

$$\sigma_z = \sqrt{\frac{\sum_{j=1}^{n}\left(z_j - z_j^p\right)^2}{n}},$$

а показатель надёжности уравнения связи

$$R = \sqrt{1 - \sum_{j=1}^{n}\left(z_j - z_j^p\right)^2 \Big/ \sum_{j=1}^{n}\left(z_j - \overline{z_j}\right)^2} \to 1,$$

где $\bar{z} = \sum_{j=1}^{n} z_j \Big/ n$ характеризует тесноту связи результирующего показателя с выбранными факторами, т.е. правильность выбора вида формальной зависимости.

В идеале ошибка уравнения связи является независимой случайной величиной с нормальным распределением. Если математическое ожидание ошибки равно нулю при постоянном и конечном значении дисперсии, то можно считать, что полученные параметры выбранной зависимости, рассчитанные методом наименьших квадратов, не смещены по величине и, следовательно, полученная факторная модель достаточно точно и надёжно отражает реально существующую зависимость.

Однако далеко не всегда ошибка отвечает указанным условиям и тогда, вообще говоря, метод наименьших квадратов даёт смещенные оценки параметров. Одним из наиболее характерных отклонений «от нормы» является случай автокорреляции ошибок.

Ряд значений переменной величины $\Delta_1, \Delta_2, ..., \Delta_k$ называется автокоррелированным, если имеет место корреляционная связь между последовательными значениями переменной в этом ряду. Особенно часто автокорреляция наблюдается во временных рядах, когда последующие уровни рассматриваемого показателя оказываются корреляционно связанными с предыдущими уровнями. Один из простых способов проверки автокорреляции заключается в определении и оценке значимости коэффициентов корреляции между сдвинутыми на одно значение переменной рядами, т.е. между рядами

$$\Delta_1, \Delta_2, ..., \Delta_{k-1} \quad \text{и} \quad \Delta_2, \Delta_3, ..., \Delta_k .$$

Основными причинами появления автокорреляции ошибок управления являются:

1) неадекватность по отношению существующей зависимости избранной формы уравнения связи, когда, например нелинейную зависимость пытаются отобразить с помощью линейного уравнения регрессии;

2) неполнота охвата показателей факторов, когда за пределами уравнения связи остался один или несколько факторов, оказывающих существенное влияние на величину зависимости переменной.

Если ошибки уравнения связи автокоррелированы, то для получения более надёжного уравнения с несмещёнными параметрами необходимо устранить или уменьшить автокорреляцию путём улучшения формы уравнения связи и более глубокого анализа показателей факторов с тем, чтобы важнейшие из них были явно представлены в факторной модели.

Отклонение параметров уравнения связи и коэффициентов корреляции, связанные с неполярным охватом всей генеральной совокупности однородных в качественном отношении объектов, представляют собой ошибки выборки.

Анализ ошибок выборки необходим для определения существенности (значимости) найденных характеристик факторной модели (коэффициентов корреляции и регрессии).

В результате анализа существенности коэффициента корреляции выбранной совокупности определяется, может ли найденная величина данного коэффициента обуславливаться случайными колебаниями или она для этого слишком велика и, следовательно, с большей вероятностью свидетельствует о наличии корреляционной связи и в генеральной совокупности.

Анализ существенности коэффициентов регрессии (параметров факторной модели) позволяет судить о значимости включенных в уравнение связи переменных.

Если анализ показывает, что величина выборочного коэффициента регрессии несущественно в статистическом отношении отличается от нуля, то следует рассмотреть вопрос об исключении из уравнения связи соответствующей переменной.

Наиболее простой практический приём заключается в исключения из уравнения связи

$$\mathbf{z = a_0 + \sum_{i=1}^{m} a_i u_i} \, ,$$

соответствующей переменной $\mathbf{u_j}$, $\mathbf{j \in \{\overline{1, m}\}}$, когда средняя квадратическая ошибка соответствующего коэффициента регрессии $\mathbf{a_j}$, которая может быть приближённо оценена, как

$$\mathbf{\sigma_{a_j} = \frac{\sigma_z}{\sigma_{u_j}} \times \frac{1 - r^2_{zu_j}}{\sqrt{n}}}$$

превышает его абсолютную величину

$$\mathbf{\sigma_{a_j} \geq |a_j|} \, .$$

Таким образом, анализ существенности коэффициентов регрессии способствует улучшению формы уравнения факторной модели путём сохранения в нём лишь статистически значимых факторов.

Ошибки выборочных характеристик факторной модели позволяют не только провести анализ их существенности, но и определить для статистически значимых характеристик доверительные пределы, т.е. минимальную и максимальную границы, в которых с заданной доверительной вероятностью заключена величина показателя в генеральной совокупности.

Общие методы, применяемые в математической статистике для анализа существенности параметров и определения доверительных границ, могут дать искаженные результаты при наличии автокорреляции одной или нескольких переменных, входящих в уравнение связи факторной модели.

Автокорреляция переменных представляет собой весьма распространённое явление при построении факторных моделей на основе временных рядов исходных данных, для которых зависимость последующих значений признака от его предыдущих значений является достаточно характерной.

Проверка наличия автокорреляции переменных производится на основе расчёта коэффициентов автокорреляции между исходным рядом данных и рядами, сдвинутыми на один или несколько интервалов наблюдения.

Добиться устранения автокорреляции, как правило, удаётся путём исключения из временного ряда линейной или нелинейной систематической тенденции изменения показателя во времени. Практически для исключения линейной тенденции достаточно образовать исходный ряд данных не из абсолютных значений признака $\left(u_{j1}, \ldots, u_{jn}\right)$, а из последовательных, так называемых, первых разностей $\left(u_{j2} - u_{j1}\right), \left(u_{j3} - u_{j2}\right), \ldots, \left(u_{jn} - u_{j(n-1)}\right)$.

В случае нелинейных тенденций берутся разности более высоких порядков – вторые, третьи и т.д.

Вместо непосредственного устранения линейной тенденции можно применять следующий приём: вводить время в качестве самостоятельной независимой факторной модели

$$k_0 = A \prod_{i=1}^{m} k_i^{a_i} e^{a_{m+1} t} .$$

Наблюдаемые статистические значения независимых переменных, служащие основой для вычисления параметров факторной модели, обычно содержат ошибки наблюдения, которые обуславливаются следующими основными причинами:

1) необходимость количественного выражения факторов, которые, во-первых, являются качественно неоднородными в силу своей высокой агрегированности и, во-вторых, не остаются качественно неизменными даже на протяжении сравнительно коротких промежутков времени;

2) наблюдаемые статистические данные, представляя собой, как правило, определённый вид отчётности, могут оказаться искажёнными в интересах отчитывающегося.

Если есть основания предполагать, что ошибки наблюдения не очень велики (порядка нескольких процентов), то для расчёта параметров факторной модели применяют метод наименьших квадратов.

При значительной величине ошибок наблюдения метод наименьших квадратов может дать сильно смещённые значения параметров. В этом случае требуется применение другого метода оценки параметров, например метода взвешенной регрессии.

В связи с ошибками наблюдений серьёзное значение приобретает исследование мультиколлинеарности. Мультиколлинеарность проявляется в том, что наряду с изучаемой корреляционной связью - между зависимой переменной (z) и независимыми величинами $\left(u_1, \ldots, u_m\right)$ - в исследуемой совокупности существуют и другие корреляционные связи - между самими независимыми переменными $\left(u_1, \ldots, u_m\right)$. Мультиколлинеарность «опасна» тем, что полученные расчётным путём параметры факторной модели могут оказаться бессодержательными, обусловленными не существенными отношениями исследуемой зависимости, а ошибками наблюдения.

Простейший способ проверки мультиколлинеарности заключается в вычислении и оценке величины парных коэффициентов корреляции $\left(r_{u_i u_j}\right)$ для каждой пары включаемых в уравнение связи независимых переменных $\left(u_i, u_j\right), i \in \left\{\overline{1,m}\right\}, j \in \left\{\overline{1,m}\right\}$. Если для какой-либо пары переменных $\left(u_i, u_j\right)$ парный коэффициент корреляции $\left(r_{u_i u_j}\right)$ оказывается по абсолютной величине больше 0,8, что, как было отмечено выше, свидетельствует о тесной корреляционной связи между ними, то во избежание получения бессодержательных коэффициентов регрессии следует исключить из уравнения связи одну из этих переменных.

Однако исключение из факторной модели

$$k_0 = A\prod_{i=1}^{m} k_i^{a_i}$$

одной из взаимосвязанных переменных не является строго обязательным условием и не применяется в тех случаях, когда каждый из взаимосвязанных факторов $\left(k_i, k_j\right)$ оказывает на результирующий показатель k_0 достаточно сильное специфическое воздействие. Это воздействие количественно оценивается найденными коэффициентами эластичности $\left(a_i\right), i = \left\{\overline{1,m}\right\}$ результирующего показателя k_0 по рассматриваемым факторам $\left(k_i\right), i = \left\{\overline{1,m}\right\}$, которые численно показывают, на сколько процентов изменится результирующий показатель k_0, если характеризующий фактор $\left(k_i\right), i = \left\{\overline{1,m}\right\}$ изменится на один процент при неизменных значениях всех остальных факторов.

Таким образом, во избежание получения искаженных оценок параметров создаваемой факторной модели и неверных выводов значимости обуславливающих факторов следует по возможности исключать характерные для эмпирических расчетов параметров модели явления мультиколлинеарности, автокорреляции переменных и автокорреляционных ошибок уравнения.

Поступила 12.04.2002

***КУЧМИЕВ Владимир Гавриилович**, канд. техн. наук, ст. научный сотрудник, директор ЗАО "Инвест маркет Украина". В 1976 году окончил ХАИ. Область научных интересов – прогрессивные информационные технологии.*

***ЛЫСЕНКО Александр Иванович**, канд. техн. наук, доцент, доцент факультета экономики и менеджмента НАКУ "ХАИ". В 1967 году окончил ХАИ. Область научных интересов – исследование операций, теория игр. E-mail: k602@xai.edu.ua .*