МНОГОФАКТОРНЫЕ ВЗАИМОСВЯЗИ В ДЕЯТЕЛЬНОСТИ РЕГИОНАЛЬНОГО ТРАНСПОРТНОГО КОМПЛЕКСА

О.В. Буреш, канд. экон. наук, доцент ГОУ “ОГУ”, г. Оренбург

Особую актуальность приобретает изучение взаимосвязей спроса и предложений транспортного комплекса региона в условиях развивающейся рыночной экономики. Изучение механизма рыночных связей, выявление роли факторов в изменении сложных явлений, влияние объема и состава предлагаемых услуг на развитие транспорта, формирование новых услуг, прибыли и других качественных показателей имеет первостепенное значение для анализа и прогнозирования конъюнктуры рынка, рациональной организации транспортных процессов и решения многих вопросов успешного функционирования комплекса.

Формирование системы показателей и их взаимосвязей, адекватно отражающих результаты деятельности организаций, предприятий, фирм транспортного комплекса является одной из основных задач анализа и прогнозирования его развития. В условиях командно-административной экономики система показателей в основном состояла из двух подсистем, включающих в себя:

- директивные показатели, необходимые для отчета низших структур комплекса перед вышестоящими органами;
- расчетные показатели, которые использовались для внутреннего экономического анализа структур транспортного комплекса и определялись директивными показателями.

В зависимости от директивных показателей, от проводимых экономических реформ изменялся круг и состав расчетных показателей. Так, в период экономической реформы 1965 г. он сузился, а во времена централизованных начал управления экономикой, наоборот, расширился. Начиная с 1965 г. впервые в систему показателей были введены прибыль и рентабельность.

С начала периода перестройки командно-административной экономики в рыночную, полностью исчезли директивные показатели, и всем предприятиям, в том числе организациям, предприятиям и фирмам транспортного комплекса была предоставлена полная свобода в выборе экономической стратегии. Однако, исходя из сложившейся общеэкономической ситуации в стране, вопросы внутреннего планирования, анализа динамики развития организаций, предприятий, фирм комплекса, отодвинулись на задний план, главным стала проблема их выживания. Постоянно изменяющиеся экономические условия, такие как инфляция, спад производства, снижение спроса и др., ввело организации, предприятия, фирмы комплекса в кризисное и предбанкротское состояние, что требовало пересмотра системы показателей, отражающих рыночное существование транспортного комплекса и его структур.

На сегодняшний день проблема создания системы показателей, отображающих всесторонние аспекты рыночного существования транспортного комплекса региона и его организаций, предприятий, фирм далека от своего логического завершения. Системы экономических показателей, в настоящее время, отличаются своей односторонностью, представляют собой набор отдельных экономических характеристик, не являются комплексными и отстают от требований сегодняшнего времени.

В условиях становления рыночных отношений система показателей развития транспортного комплекса должна ориентироваться на рынок и отражать сложившиеся условия, в которых функционирует комплекс, экономическое состояние организаций, предприятий, фирм комплекса, их взаимосвязь с рынком и результаты деятельности за рассматриваемый период.

С другой стороны, при научно обоснованном исследовании и рациональном управлении механизмом рыночных отношений важно не только создание оптимальной системы показателей, но и изучение взаимосвязи показателей деятельности транспортного комплекса, необходимой не только для установления такового факта, но и для придания выявленным связям математической сути. Так как в практической деятельности невозможно использовать результаты экономических исследований закономерностей связи без их количественной оценки.

При математическом моделировании развития регионального транспорта учесть влияние различных факторов возможно с помощью многофакторных эконометрических моделей, которые позволяют достаточно полно отобразить многообразие конкретной деятельности комплекса, отобразить существующие взаимосвязи внутри системы и с окружающей средой, определить степень воздействия отдельных факторов на исследуемую систему.

Математическую платформу эконометрических моделей, основанных, в определенной степени, на инерционности взаимосвязей, составляют методы корреляционного, факторного и регрессионного анализа.

В процессе решения широкого круга задач деятельности транспортного комплекса, прежде всего, необходимо знать связаны или не связаны между собой две или более случайные величины.

Решение такого рода задач обычно сводится к выявлению зависимости результативного признака y от факторного признака x, которая проявляется частично, т. к. возможно влияние других различных факторов e:

$$y = y\left(x\right) + e .$$

Такие зависимости являются соотносительными не жесткими. При изучении корреляционной связи исследуется влияние учтенных факторных признаков при абстрагировании от других, что позволяет устанавливать закономерности взаимодействия анализируемых показателей и получать количественные характеристики корреляционной связи.

Таким образом, существование взаимосвязей и их тесноту можно выразить коэффициентом корреляции r. Предположим, что производится выборка из случайных величин x_i и x_j, в результате чего получается N пар их наблюдаемых значений. Коэффициент парной корреляции можно оценить по этим выборочным значениям, следующим образом:

$$r_{i,j} = \frac{\sum_{i,j=1}^{N}(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x}_i)^2 \sum_{j=1}^{N}(x_j - \bar{x}_j)^2}}.$$

Как и величина r, коэффициент парной корреляции $r_{i,j}$ лежит в пределах $-1 \le r_{i,j} \le +1$ и является количественной мерой линейной связи между показателями, отрицательное значение указывает на обратную связь между величинами. Чем больше абсолютное значение коэффициента корреляции между зависимым и независимым признаком, тем в большей степени зависимый признак определяется данным независимым. Граничные значения достигаются только в том случае, когда наблюдения обнаруживают идеальную линейную зависимость. Если же зависимость отлична от линейной и (или) наблюдается разброс исследуемых значений, то независимо от того, обусловлено ли это обстоятельство ошибками измерений или нелинейным характером связи анализируемых величин, коэффициент уменьшается.

Для проверки значимости коэффициента корреляции необходимо определить $t-$ распределение с $N-2$ степенями свободы [1]:

$$t_r = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}},$$

где $r-$ выборочный условный коэффициент корреляции.

Воспользовавшись данной формулой можно получить границы для r, зная границы для t_r, соответствующие определенному уровню значимости. Границы для r табулированы в зависимости от числа степеней свободы $N-2$ и уровня значимости a. Если $|r_{набл}| \le r_{табл}$, то гипотеза равенства нулю коэффициента корреляции, т. е. $r_{i,j} = 0$ не отвергается с вероятностью ошибки a, в противном случае отвергается.

Корреляционный анализ можно углубить. Для нахождения интервальной оценки коэффициента корреляции $r_{i,j}$ целесообразно ввести в рассмотрение логарифмическую функцию Фишера:

$$Z = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right).$$

Как известно, случайная величина Z распределена приблизительно нормально, в то время как выборочный коэффициент корреляции r распределен крайне неравномерно. Средняя квадратическая ошибка случайной величины Z определяется как:

$$s_Z = \frac{1}{\sqrt{N-3}}.$$

Откуда видно, что распределение случайной величины Z не зависит от степени тесноты связи, т. к. средняя квадратическая ошибка определяется только объемом случайной выборки N.

Генеральное среднее распределения:

$$MZ \approx \frac{1}{2} ln\frac{1+r_{ij}}{1-r_{ij}} + \frac{r_{ij}}{2(N-1)}.$$

В таком случае доверительный интервал, оценивающий MZ, с надежностью $g = 1 - a$ будет иметь следующий вид:

$$Z - t_g \sqrt{\frac{1}{N-3}} \le MZ \le Z + t_g \sqrt{\frac{1}{N-3}},$$

где t_g – определяется по таблицам интеграла Лапласа.

По таблицам Фишера и Иейтсома можно перейти от Z к $r_{i,j}$:

$$r_{min} \le r_{ij} \le r_{max},$$

где r_{min} и r_{max} выбирают в зависимости от Z, пренебрегая поправочным членом $\frac{r_{ij}}{2(N-1)}$.

Проверку результатов корреляционного анализа возможно осуществлять с использованием $F-$критерия, разработанного английским статистиком Р. Фишером, т. к. анализ корреляции тесно связан с дисперсионным анализом.

С помощью корреляционного анализа, возможно, установить на сколько тесна связь между двумя или более случайными величинами. Однако в дополнение к этому желательно располагать моделью зависимости, которая позволяла бы определить значение некоторой величины (зависимой) по заданным значениям других величин (независимых). Для этих целей достаточно широко используется регрессионный анализ, позволяющий получать зависимости результативного признака от факторных признаков – однофакторные и многофакторные.

В процессе получения однофакторных моделей и при переходе к многофакторным, необходимо выявить факторы, наиболее существенно влияющие на результирующий признак. Для решения этой задачи существует несколько методов, среди которых методы исключения и включения, метод итераций. Метод исключения состоит в том, что на первом этапе строится регрессионная модель на основе всех факторных признаков, а затем сокращают число факторов до тех пор, пока не будет достигнуто заранее заданное условие. Метод включения заключается в последовательном включении факторов в модель, до тех пор, пока не получится модель с заранее заданным уровнем качества. Методы исключения и включения имеют недостаток, заключающийся в том, что при наличии корреляции между независимыми признаками изменение критериальных параметров, таких как коэффициент множественной корреляции, остаточная дисперсия и др., зависит еще и от порядка исключения или введения независимых признаков. Метод итераций, основанный на сравнении коэффициентов корреляции, вычисленных относительно остаточной компоненты результирующего фактора и факторных признаков, является наиболее приемлемым.

Для определения существования взаимозависимостей между показателями экономических процессов, возможно, использовать факторный анализ – метод главных компонент [2, 3].

Метод главных компонент совпадает с методом расчленения ковариационной или корреляционной матрицы на совокупность ортогональных векторов (компонент) или направлений по числу рассматриваемых переменных.

Указанные векторы соответствуют собственным векторам и собственным значениям корреляционной матрицы. По этому методу собственные значения выделяются в порядке убывания их величины, что становится существенным, если для описания данных должно быть использовано лишь незначительное число компонент.

Векторы попарно ортогональны, и компоненты, полученные по ним, некоррелированы. Хотя несколько компонент могут выделить большую часть суммарной дисперсии переменных, однако для точного воспроизведения корреляций между переменными требуются все компоненты.

В тех случаях, когда применяется метод главных компонент, не нужно делать никаких гипотез о переменных, они не обязаны даже быть случайными величинами.

Метод главных факторов, центроидный метод в противоположность методу главных компонент заранее объясняет матрицу ковариаций наличием минимального или, по крайней мере, небольшого числа гипотетических переменных или факторов.

В то время, как метод главных компонент ориентирован на дисперсии, метод главных факторов, центроидный метод ориентированы на ковариации (или на корреляционную связь), в которых основным предположением является равенство:

$$x_{\mathrm{i}} = \sum_{\mathrm{r}=1}^{\mathrm{k}} a_{\mathrm{ir}} F_{\mathrm{r}} + l_{\mathrm{i}}, (\mathrm{i} = 1,2,\ldots,p), \quad (1)$$

где x_i – i – ая переменная;

F_{r} – r – ый фактор;

a_{ir} – факторная нагрузка;

k – количество факторов (точно задано);

l_i – остатки, которые представляют источники отклонений, действующие только на x_i.

Все p случайных величин l_i предполагаются независимыми как между собой, так и с k величинами F_r.

Уравнения вида (1) нельзя проверить непосредственно, поскольку p переменных x_i выражены в них через $(p+k)$ ненаблюдаемых переменных. Однако эти уравнения заключают в себе гипотезу о ковариациях и дисперсиях x_i, которую можно проверить.

Когда число факторов $k>1$, то ни факторы, ни нагрузки не определяются однозначно, поскольку в уравнении (1) факторы F_r могут быть заменены любым ортогональным преобразованием нагрузок. Это свойство использовано для преобразования или вращения факторов, полученных в каком-либо практическом исследовании.

Вращение подбирается так, чтобы переменные, которые в большей или меньшей степени измеряют некоторые легко опознаваемые стороны, имели бы достаточно высокие нагрузки на один фактор и нулевые или почти нулевые на другие факторы.

Если отправной точкой является корреляционная матрица R с единицами на главной диагонали, то говорят о методе главных компонент, чья модель отлична от моделей метода главных факторов, центроидного и приводит к дескриптивным факторам. Если в матрице R используют оценки общностей, то получают модель метода главных факторов или центроидного.

Классическая модель факторного анализа имеет вид:

$R = A \cdot C \cdot A'$, (2)

где R – корреляционная матрица;

A – матрица факторных нагрузок;

C – корреляционная матрица, отражающая связи между факторами;

A' – транспонированная матрица факторных нагрузок.

Если наложить условие некоррелированности факторов, т.е. $C = I$, где I – единичная матрица, то в результате получим:

$R = A \cdot A'$. (3)

Система уравнений, соответствующая (3) имеет однозначное решение с вводом дополнительных условий, а именно: сумма квадратов нагрузок первого фактора должна составлять максимум от полной дисперсии; сумма квадратов нагрузок второго фактора должна составлять максимум оставшейся дисперсии и т.д., т.е. максимизирует функцию:

$$\mathrm{S}_1 = \sum_{\mathrm{i}=1}^{\mathrm{m}} a_{\mathrm{i}1}^2 = max,$$

при $\frac{m(m-1)}{2}$ независимых друг от друга условиях.

$$r_{\mathrm{ik}} = a_{\mathrm{i}1} a_{\mathrm{k}1}, (i,k = 1,2,\ldots,m, i<k),$$

где m – число переменных в матрице наблюдений.

Для максимизирования функции, связанной некоторым количеством дополнительных условий, используем метод множителей Лагранжа.

В результате приходим к системе m однородных уравнений с m неизвестными a_{i1}:

$$\begin{cases}(1-l)a_{11}+r_{12}a_{21}+...+r_{1m}a_{m1}=0\\ r_{21}a_{11}+(1-l)a_{21}+...+r_{2m}a_{m1}=0\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\\ r_{m1}a_{11}+r_{m2}a_{21}+...+(1-l)a_{m1}=0\end{cases} \quad (4)$$

Система равенств (4) составляет проблему собственных значений действительной симметричной матрицы. В общем она записывается в следующим образом:

$Ra_l = l_l a_l$ или $(R-l_l)a_l = 0$,

где l_l – собственные значения, они соответствуют собственным векторам a_l матрицы R.

Факторы пропорциональны собственным векторам матрицы R.

Путем нормирования получим искомое значение a_{il} матрицы A по компонентам собственных векторов матрицы R:

$a_{il} = a_{il}\sqrt{l_l}\Big/\sqrt{a_{1l}^2 + a_{2l}^2 + ... + a_{ml}^2}$.

Рассмотренный метод главных компонент не дает точного ответа на вопрос о том, сколько же факторов определяют взаимодействие переменных в генеральной совокупности.

Эту проблему можно решить, используя метод максимального правдоподобия. В этом случае предполагается, что наблюдаемые переменные нормально распределены, а общие факторы ортогональны.

С учетом характерных факторов уравнения (3) будет иметь следующий вид:

$R = AA' + U^2$.

Для определения максимально правдоподобных оценок a_{1l} и U_i^2 максимизируется функция правдоподобия. Это приводит к большому числу возможных результатов. Из них выбирается тот, который удовлетворяет следующему условию, что:

$J = A'\left[U^2\right]^{-1}A$, (5)

является диагональной матрицей. Равенство (5) приводится к следующему виду:

$A = J^{-1}A^1\left[U^2\right]^{-1}\left(R - U^2\right)$. (6)

Кроме этого, имеет место равенство:

$U_i^2 = 1 - \sum_{l=1}^{r} a_{il}^2$. (7)

Исходя из более или менее произвольно взятых первых приближений для A и U^2 в правой части (6), в левой части этого равенства получают новую матрицу A, а из (7) - новую матрицу U^2, которые можно рассматривать, как хорошие приближения к истинным матрицам.

Далее итеративная процедура повторяется, в результате которой все полученные матрицы снова подставляются в формулы (6) и (7).

Процедура сходится очень медленно, и бывают случаи, когда достигаются малые разности между последовательными итерациями, но приближения еще находятся далеко от истинных значений. Кроме того, имеются корреляционные матрицы, для которых итерационный процесс не сходится.

Итеративную процедуру решения уравнения (6) можно рассматривать в качестве особого способа определения собственных значений и собственных векторов матрицы $\left[U^2\right]^{-1}\left(R-U^2\right)$, причем A^1 содержит собственные векторы, а диагональная матрица J^{-1} собственные значения этой матрицы.

Тест проверки значимости числа факторов, найденного методом максимального правдоподобия:

$$\aleph^2 = (n-1)ln\frac{\left|R^+\right|}{\left|R\right|} .$$

В этой формуле $\left|R^+\right|$ – определитель матрицы корреляций, воспроизведенных с помощью выбранной модели, $\left|R\right|$ – определитель исходной корреляционной матрицы, n – число наблюдений.

Величина $\aleph^2$ имеет приблизительно $\aleph^2$ распределение с $\frac{1}{2}\left[(m-r)^2-m-r\right]$ степенями свободы. Здесь m – число переменных, r – число выделенных факторов.

Если при определенном r вычисленное значение критерия превышает табличное значение $\aleph^2$, соответствующее заданному уровню значимости, это указывает на то, что необходимо выделить больше факторов, чем r, по крайней мере $r+1$.

После определения матрицы факторных нагрузок A для лучшей интерпретации факторов используют вращение A в пространстве общих факторов. В настоящее время наиболее распространен "метод варимакс" для осуществления вращения матрицы A. Метод варимакс был предложен Кайзером [4]. Согласно Кайзеру, простота фактора определяется дисперсией квадратов его нагрузок. Если эта дисперсия максимальна, то отдельные его нагрузки близки к нулю или единице, т.е. он описывается наиболее просто и поэтому его можно наилучшим образом проинтерпретировать.

Дисперсия квадратов нагрузок фактора l равна:

$$S_l^2 = \frac{1}{m}\sum_{i=1}^{m}\left(B_{il}^2\right)^2 - \frac{1}{m^2}\left[\sum_{i=1}^{m}B_{il}^2\right]^2 .$$

Просуммировав дисперсию по всем факторам, получим величину, которая будет максимальной тогда, когда дисперсия квадратов нагрузок каждого фактора примет наибольшее значение:

$$\sum_{l=1}^{r}S_l^2 = \frac{1}{m}\sum_{l=1}^{r}\sum_{i=1}^{m}B_{il}^4 - \frac{1}{m^2}\sum_{l=1}^{r}\left[\sum_{i=1}^{m}B_{il}^2\right]^2 = max. \quad (8)$$

Данный критерий имеет недостаток, который заключается в том, что переменная с большей общностью сильнее влияет на значение угла поворота, чем переменная с меньшей общностью, т.е. она обладает большим весом при определении финального решения.

Кайзер предложил делить факторные нагрузки на соответствующие общности, благодаря чему все факторы-переменные приводятся к длине, равной единице.

Таким образом, при определении положения осей координат имеют дело с нормированными переменными с равными весами. В отличие от (8) модифицированный Кайзером варимакс - критерий умножается еще на m:

$$m\sum_{l=1}^{r}\sum_{l=1}^{m}\left(B_{il}/h_{i}\right)^{4}-\sum_{l=1}^{r}\left[\sum_{l=1}^{m}B_{il}^{2}/h_{i}^{2}\right]^{2}=max.$$

Нахождение максимума функции приводит к определению положения системы координат, которое удовлетворяет требованиям ортогональной простой структуры.

Достаточная простота и хорошая программная реализация метода главных компонент позволили сократить время машинной обработки и анализа полученных результатов по сравнению с обычной схемой нахождения интересующих связей между показателями экономического процесса.

Однако довольно часто при исследовании показателей экономического процесса, полученная матрица факторных нагрузок в алгоритме метода главных компонент получается вырожденной и тогда невозможно осуществить ортогональное вращение матрицы факторных нагрузок. Поэтому предварительно ее необходимо уплотнить. Предлагается алгоритм уплотнения, который состоит из следующих шагов:

1) проверяем матрицу факторных нагрузок, которая задана векторно, на вырожденность;
2) создаем вектор, содержащий номера нулевых строк;
3) если в полученном векторе номер нулевой строки равен числу признаков, то необходимо уменьшить число строк на 1 и перейти к пункту 8;
4) создаем вектор, содержащий элементы следующей после нулевой строки;
5) помещаем в исходной матрице факторных нагрузок на место нулевой строки элементы созданного вектора;
6) переходим к дальнейшему просмотру вектора, содержащего номера нулевых строк и повторяем с пункта 4;
7) уменьшаем величину переменной, учитывающую число параметров, на количество нулевых строк;
8) осуществляем сжатие матрицы факторных нагрузок.

После ортогонального вращения матрицы факторных нагрузок, для облегчения анализа полученных результатов, необходимо осуществить объединение признаковов, по максимальной нагрузке, в каждом факторе.

Предлагается следующий алгоритм объединения признаков:

1) создаем массив результирующих номеров строк, принимая во внимание возможность вырождения матрицы факторных нагрузок перед варимаксным вращением;
2) находим максимальные, по абсолютной величине, нагрузки для каждого показателя в матрице факторных нагрузок после варимаксного вращения;
3) определяем, какому фактору соответствует максимальная нагрузка показателя;
4) находим объединенные признаки в каждом факторе.

Таким образом, методика выявления связей при анализе функционирования сложной экономической системы, к которой принадлежит региональный транспортный комплекс, состоит из следующих этапов:

1) расчет корреляционной матрицы;
2) вычисление собственных значений, собственных векторов корреляционной матрицы;
3) определение накопленных отношений собственных значений корреляционной матрицы, больших или равных заданной пользователем константы;
4) вычисление матрицы факторных нагрузок по собственным значениям и соответствующим собственным векторам корреляционной матрицы;
5) уплотнение матрицы факторных нагрузок, если она вырождена;
6) ортогональное вращение матрицы факторов;
7) определение признаков, объединившихся в каждом факторе.

Таким образом, для выявления существования зависимостей между показателями предлагается использовать метод главных компонент. Современные ППП не позволяют получать промежуточные результаты, обрабатывать варожденные матрицы и не выдают конечный результат в виде удобном для проведения необходимого анализа, поэтому программно реализован алгоритм метода главных компонент с учетом симметрии корреляционной матрицы, который дополнен алгоритмом уплотнения матрицы факторных нагрузок для случаев ее вырождения и алгоритмом определения связей по матрице факторных нагрузок после варимаксного вращения. Программный модуль может представлять собой отдельную компоненту в программном комплексе исследования регионального транспорта, созданном на основе модели компонентных объектов.

ЛИТЕРАТУРА

1. Дубров А. М., Мхитарян В. С., Трошин Л. И. Многомерные статистические методы. – М.: Финансы и статистика, 1998.
2. Иберла К. Факторный анализ. -М.: Статистика, 1980.
3. Харман Г. Современный факторный анализ. -М.: Статистика,

1972.
4. Kaiser.H.F. [1]. The varimax criterion for analytic rotation in factor analysis. Psychometrica, 23, 187 - 200(1958).