

УДК 519.862.6

АНАЛИТИЧЕСКИЙ АППАРАТ МОДЕЛИРОВАНИЯ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫХ ЗАВИСИМОСТЕЙ

Гусарова О.М.

*ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации», Смоленск,
e-mail: om.gusarova@mail.ru*

Осуществлено исследование основных проблем моделирования корреляционно-регрессионных зависимостей экономико-статистических данных. Выявлены проблемы и этапы реализации спецификации эконометрических моделей. Проведен анализ математического аппарата выявления корреляционной взаимосвязи статистических данных. Осуществлено исследование проблем выбора факторных признаков для построения модели множественной регрессии. Выявлены особенности оценки качества регрессионных моделей парной и множественной регрессии. Дана краткая характеристика и анализ особенностей использования различных критериев для оценки статистической значимости экзогенных переменных при осуществлении эконометрического моделирования. Выполнен обзор современных информационно-аналитических технологий анализа и прогнозирования экономико-статистических данных. Осуществлен анализ факторов, определяющих достоверность интервального прогноза экономико-статистических показателей.

Ключевые слова: корреляционно-регрессионная зависимость, модель множественной регрессии, анализ качества модели, построение интервального прогноза

ANALYTICAL APPARATUS-CORRELATION REGRESSION MODELING DEPENDENCIES

Gusarova O.M.

*Financial University under the Government of the Russian Federation, Smolensk,
e-mail: om.gusarova@mail.ru*

Carried out a study of basic problems of modelling correlation and regression dependences of Economics and statistics. The problems and the steps to implement the specification of econometric models. Analysis of the mathematical device to identify the correlation relationship statistics. Carried out a study of the problems of choice factor signs to build multiple regression models. Peculiarities of quality assessment regression models steamy and multiple regression. Brief description and analysis of the use of different criteria for assessing the statistical significance of exogenous variables in the implementation of econometric modelling. Review of modern analytical technologies for analyzing and forecasting economic and statistical data. Analyses the factors determining validity interval prediction of economic and statistical indicators.

Keywords: correlation and regression, multiple regression model dependence, quality analysis, model building interval prediction

В условиях интенсивного внедрения в практическую деятельность инноваций одним из перспективных направлений повышения эффективности управленческой деятельности является использование эконометрического моделирования, позволяющего осуществить количественный и качественный анализ взаимосвязанных финансово-экономических показателей [1, 3]. Корреляционно-регрессионная зависимость экономических данных, имеющая место в силу причинно-следственных отношений, позволяет выявить и обосновать ряд характеристик, способствующих повышению эффективности управленческих решений.

Эконометрическое моделирование, базирующееся на корреляционно-регрессионной зависимости признаков, позволяет не только выявить и формализовать форму связи между разнородными экономическими показателями, но и осуществить прогноз дальнейшего развития объекта (системы, процесса) исследования [2].

Одним из основных направлений эконометрического моделирования является построение моделей парной и множественной регрессии, обладающих высоким качеством, с целью дальнейшего их использования для анализа и прогнозирования динамики экономических процессов в условиях экономической неопределенности [4]. Проблема построения регрессионных моделей высокого качества в значительной степени зависит от правильной спецификации модели, которая в дальнейшем определяет все свойства теоретической модели и возможность ее использования для построения прогнозов. Как правило, спецификация регрессионной модели предусматривает решение нескольких проблем: во-первых, анализ исходных статистических данных с целью выбора и обоснования результативного и факторных признаков; во-вторых, определение вида регрессионной модели, в максимальной степени отражающей реальные свой-

ства совокупности статистических данных, подлежащих исследованию.

Эконометрические модели, базирующиеся на исследовании корреляционно-регрессионной зависимости экономико-статистических данных, находят широкое применение в экономических и технико-экономических исследованиях, посвященных проблемам эффективного менеджмента [7]. Каждой области экономических исследований, связанной с анализом эмпирических данных, как правило, соответствуют свои эконометрические модели [6].

Определение результативного и группы факторных признаков, участвующих в построении регрессионной модели, целесообразно осуществить в два этапа: первоначально эндогенная (зависимая) и экзогенные (независимые) переменные выделяются на основании экономического анализа существа проблемы, сопряженной с целью исследования; затем дальнейшее обоснование выбора результативного и факторных признаков осуществляется, используя математический аппарат, в частности, экономико-математический анализ матрицы парных корреляций. Элементы матрицы парной корреляций являются коэффициентами парных корреляций, характеризующих тесноту взаимной связи признаков, и определяются следующим соотношением

$$\begin{aligned} r_{x,y} &= \frac{\text{Cov}(X, Y)}{S_x \cdot S_y} = \\ &= \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{S_x \cdot S_y} = \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \end{aligned} \quad (1)$$

где y , x_i – значения признаков, состоящих в корреляционной зависимости, S_x и S_y – оценка дисперсий признаков, характеризующих вариабельность этих переменных, т.е. степень их разброса относительно среднего значения.

Оценка дисперсии признаков может быть выполнена следующим образом

$$\begin{aligned} S_x &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \\ S_y &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}. \end{aligned} \quad (2)$$

где n – длина временного ряда исследуемых признаков.

Коэффициент парной корреляции колеблется в пределах от -1 до $+1$, чем ближе его значение, взятое по модулю, к единице, тем более тесной признается корреляционная взаимосвязь между признаками. Положительная или отрицательная корреляция между признаками находит свое отражение в знаке коэффициента парной корреляции.

Величина коэффициента парной корреляции сама по себе еще не свидетельствует о статистической значимости влияния выявленного факторного признака на резуль- тативный признак. Для уточнения данного момента используется t-критерий Стьюдента, который может быть определен следующим образом

$$t = \sqrt{\frac{r_{yx}^2}{1 - r_{yx}^2}} (n - 2). \quad (3)$$

При сравнении рассчитанного значения критерия Стьюдента с табличной величиной можно судить о статистически значимой корреляционной связи между признаками.

Для подтверждения правильности выбора группы факторных признаков для построения регрессионной модели целесообразно оценить коэффициент множественной корреляции, характеризующий влияние множества факторных признаков на резуль- тативный признак

$$R = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^n \varepsilon_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (4)$$

Коэффициент множественной корреляции может быть рассчитан для различного набора факторных признаков и, таким образом, можно получить линейку величин для уточнения корреляционно-регрессионной зависимости и набора факторных признаков.

Построение регрессионной модели является творческим процессом, опирающимся на серьезные знания в области экономической теории, макро- и микроэкономики, эконометрического анализа и должно быть индивидуально в каждой конкретной ситуации [5].

После определения результативного и факторных признаков необходимо перейти ко второму этапу спецификации регрессионной модели, т.е. к выбору вида модели. В зависимости от целей и задач моделирования, а также практических аспектов прикладного исследования предварительно может быть очерчен широкий круг регрессионных моделей, в котором могут быть

учтены линейные и нелинейные, парные и множественные корреляционно-регрессионные зависимости, подлежащие построению и дальнейшему исследованию. В предположении о наличии между исследуемыми признаками линейной зависимости общий вид регрессионной модели, содержащей произвольное количество факторных признаков, выглядит следующим образом

$$y(t) = a_0 + a_1 x_1(t) + a_2 x_2(t) + \dots + a_k x_k(t) + \varepsilon(t), \quad (5)$$

где $y(t)$ – эндогенная переменная, $x_i(t)$ – экзогенные переменные, $\varepsilon(t)$ – случайная компонента, характеризующая стохастическую природу экономических данных.

В случае выбора в качестве факторного признака только одного показателя можно получить частный случай модели множественной регрессии – модель парной регрессии.

Кроме линейной зависимости между признаками может иметь место и нелинейная взаимосвязь. В этом случае целесообразно построить комплекс моделей нелинейного вида. Окончательный выбор того или иного вида модели базируется на анализе показателей качества регрессионной модели [8]. В первую очередь, необходимо выполнить расчет и анализ коэффициента детерминации, характеризующего общее качество построенного уравнения корреляционно-регрессионной зависимости. Для расчетов может быть использована следующая формула

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n \varepsilon_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (6)$$

Коэффициент детерминации характеризует долю дисперсии результативного признака, учтенного в модели регрессии с определенным набором факторных признаков. Величина коэффициента детерминации варьирует в пределах от 0 до 1 и чем ближе ее значение к единице, тем большее значение вариации результативного признака обусловлено вариацией факторных признаков, включенных в модель регрессии. Между коэффициентом детерминации и коэффициентом множественной корреляции существует однозначная зависимость

$$R = \sqrt{R^2}. \quad (7)$$

После оценки общего качества уравнения регрессии, характеризуемого коэффициентом детерминации, целесообразно

оценить статистическую значимость построенного уравнения регрессии, используя критерий Фишера, выраженный следующим соотношением

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}, \quad (8)$$

где k – количество факторных признаков, участвующих в построении модели множественной регрессии.

По итогам сравнения расчетного и табличного значения F-критерия Фишера можно сделать выводы о статистической значимости построенного уравнения регрессии [10]. Рекомендуется построить несколько видов регрессионных моделей с различным набором факторных признаков и, оценивая критерий Фишера для различных наборов факторных признаков, можно сделать вывод о наиболее приемлемом наборе экзогенных переменных, характеризующих корреляционно-регрессионную зависимость исследуемых признаков.

Для уточнения целесообразности включения в регрессионную модель конкретного факторного признака рекомендуется оценить статистическую значимость соответствующего коэффициента регрессии, характеризующего значимость влияния факторного признака на резуль- тативный признак. Критерий Стьюдента в оценке статистической значимости коэффициента регрессии при факторном признаке имеет вид

$$t_{a_i} = a_i / S_{a_i}, \quad (9)$$

где S_{a_i} – это стандартное (среднеквадратическое) отклонение коэффициента уравнения регрессии a_i . Величина S_{a_i} представляет собой квадратный корень из произведения несмещенной оценки дисперсии S_e^2 и i -го диагонального элемента матрицы, обратной матрице системы нормальных уравнений.

Если расчетное значение t -статистики Стьюдента не превосходит табличного значения критерия, то оцениваемый коэффициент регрессии и соответствующий ему факторный признак не признаются статистически значимыми, и такой факторный признак рекомендуется из регрессионной модели исключить. При этом качество построенной регрессии существенно не ухудшится. Если же расчетное значение t -статистики Стьюдента превосходит пороговое табличное значение, то коэффициент регрессии и соответствующий факторный признак признаются статистически значимыми и такой фактор целесообразно оставить в регрессионной модели.

Количество факторных признаков, составляющих модель регрессии, может быть определено исходя из следующих соображений: целесообразности при осуществлении дальнейших вычислений, например, построении прогноза, и интерпретации полученных результатов исследования. Теоретически в модель регрессии может быть включено произвольное количество экзогенных переменных, но слишком большое их количество приведет к неустойчивости модели, ее сверхчувствительности и затруднит вычислительный процесс и дальнейшую интерпретацию результатов. При использовании для анализа и прогнозирования экономико-статистических данных современных информационных технологий эта проблема может быть частично нивелирована.

Одной из основных целей моделирования корреляционно-регрессионных зависимостей экономико-статистических данных является построение прогнозов и определение перспектив развития системы (процесса). Расчет прогноза исследуемых показателей с определенной долей вероятности будет способствовать принятию правильного, своевременного, научно обоснованного управленческого решения на основании информации, полученной по регрессионной модели.

При построении точечного прогноза в уравнение регрессии необходимо подставить прогнозные значения всех экзогенных переменных, участвующих в построении регрессии. При использовании специализированных программ статистического анализа и прогнозирования экономических данных, таких как Statistica, SPSS, Vstat, Stadia, Олимп: СтатЭксперт и т.д. данная проблема существенно упрощается [9].

Для определения интервального прогноза необходимо рассчитать ширину доверительного интервала, зависящую от вероятности расчетов, заданной точности прогнозирования, длины временных рядов признаков, составляющих информационную базу исследования, количества шагов упреждения. Для расчета доверительного интервала может быть использовано следующее соотношение

$$u(X_{\text{прогн}}) = S_{\varepsilon} \cdot t_{\alpha} \cdot \sqrt{V_{\text{пр}}} = S_{\varepsilon} \cdot t_{\alpha} \cdot \sqrt{1 + X_{\text{прогн}}^T \cdot (X^T \cdot X)^{-1} \cdot X_{\text{прогн}}}, \quad (10)$$

где $X_{\text{прогн}}^T = (1, X_{1\text{прогн}}, X_{2\text{прогн}}, \dots, X_{k\text{прогн}})$,

S_{ε} – среднее квадратичное отклонение, определенное исходя из остаточной дисперсии и характеризующее точность построения прогноза,

t_{α} – статистика Стьюдента, характеризующая вероятность расчетов.

Необходимо отметить, что прогнозирование имеет вероятностный характер и отражает сложившиеся закономерности лишь при сохранении в периоде упреждения закономерностей и тенденций, имеющих место в анализируемом периоде. К тому же, прогнозирование имеет рекомендательный характер, помогающий принять правильное управленческое решение с учетом сложившихся реалий, исходя из целей деятельности и с учетом субъективных предпочтений лица, принимающего решения. Тем не менее, построение моделей корреляционно-регрессионной зависимости ряда экономико-статистических показателей и последующее определение их прогнозных значений поможет глубже понять причинно-следственные связи признаков, а, следовательно, будет способствовать принятию рационального, математически и научно обоснованного управленческого решения, и, в конечном счете, будет способствовать повышению эффективности бизнеса.

Список литературы

1. Гусарова О.М. Эконометрический анализ статистической взаимосвязи показателей социально-экономического развития России // *Фундаментальные исследования*. 2016. № 2-2. С. 357-361.
2. Гусарова О.М., Кузьменкова В.Д. Моделирование и анализ тенденций развития региональной экономики // *Фундаментальные исследования*. 2016. № 3-2. С. 354-359.
3. Гусарова О.М. Инвестиции как фактор регионального развития // *Фундаментальные исследования*. 2015. № 2-10. С. 2194-2199.
4. Голичева Н.Д., Гусарова О.М. Теория и практика моделирования финансово-экономических процессов в условиях экономической неопределенности. Смоленск: Маджента, 2016. – 236 с.
5. Гусарова О.М. Моделирование в принятии управленческих решений // *Наука и образование: проблемы и перспективы развития: сборник научных трудов по материалам Международной научно-практической конференции*. – Тамбов: Юком, 2014. – С. 41-42.
6. Гусарова О.М. Проблемы интеграции теории и практики моделирования результатов бизнеса // *Экономика и образование: Вызовы и поиск решений: сборник научных трудов по материалам II Всероссийской (заочной) научно-практической конференции (Ярославль, 15 апреля 2014 г.)* – Ярославль: Канцлер, 2014. – С. 78-82.
7. Гусарова О.М. Моделирование результатов бизнеса в менеджменте организации. // *Перспективы развития науки и образования*. – Тамбов: Бизнес-Наука-Общество, 2014. – С. 42-43.
8. Гусарова О.М. Исследование качества краткосрочных моделей прогнозирования финансово-экономических показателей. – М.: 1999. – 198 с.
9. Гусарова О.М. Информационно-аналитические технологии моделирования деятельности организаций Смоленского региона. – Смоленск: Свиток, 2013. – 100 с.
10. Орлова И.В., Половников В.А., Филонова Е.С., Гусарова О.М. и др. *Эконометрика*. – М.: Вузовский учебник, 2010. – 144 с.