Метод статистической оценки влияния факторов на взаимную торговлю на основе регрессионного анализа

Метод базируется на статистической оценке влияния независимых факторных признаков на показатели объемов взаимной торговли. Статистическая оценка воздействия факторов осуществляется на основе регрессионного анализа взаимосвязей между значениями динамических рядов результирующего и факторных признаков с использованием обобщенной модели множественной регрессии следующего вида:

$$y_t = \alpha + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_j x_{tj} + \varepsilon \quad (1)$$

И эмпирическое уравнение множественной регрессии:

$$y_t = a + b_1 x_{t1} + b_2 x_{t2} + \ldots + b_j x_{tj} + e \quad (2)$$

где: $\alpha, b_1, b_2, \ldots b_j$ - оценки теоретических значений коэффициентов обобщенной модели $\beta_1, \beta_2, \ldots \beta_j$ - или эмпирические коэффициенты регрессии, e – оценка отклонения ε.

α – свободный член модели, определяющий значение y в случае, когда все факторные переменные равны нулю.

$b_1, b_2, \ldots b_j$ – эмпирические значения коэффициентов регрессии, которые показывают абсолютное изменение во временном периоде t результата зависимой переменной y модели (2) под влиянием изменения на 1 единицу фактора x.

В данном методе изучаются взаимосвязи факторов на основе построения и диагностики линейной модели многофакторной регрессии классическим методом наименьших квадратов (МНК). Подтверждение правильности выбора функциональной формы осуществляется с помощью эконометрического теста Рамсея.

Данный тип многофакторной модели отражает условие, при котором рассматривается динамика входящих в нее переменных в каждый момент времени t. Она показывает влияние изменения независимой переменной x в некоторый момент времени t на значения зависимой переменной y. Построение эконометрической модели предполагает условие, при котором введенное препятствие приводит к изменению значений факторных переменных, отбираемых для учета влияния на объемы взаимной торговли (например, при вводе препятствия наблюдается увеличение (снижение) цен внутреннего рынка на анализируемый товар, либо объемов его производства).

Применение метода статистической оценки влияния факторов на взаимную торговлю на основе регрессионного анализа осуществляется при соблюдении ряда следующих условий:

1. Официально зафиксирован в Комиссии период введения ограничения поставок товаров во взаимной торговле между государствами – членами Союза (месяц, квартал, год).
2. Товары, подпадающие под действие препятствия, однозначно идентифицируются в соответствии кодами ТН ВЭД ЕАЭС (на уровне не выше чем 4 числовых знака кода ТН ВЭД ЕАЭС).
3. Имеются в наличии и официально опубликованы в статистических изданиях, либо на официальных сайтах уполномоченных органов или международных организаций статистические данные об объемах экспорта и (или) импорта товаров во взаимной торговле государств – членов Союза (в стоимостном или физическом выражении месячной, квартальной или годовой периодичности), а также данные о средних ценах внутреннего рынка государств-членов и средних значениях курсов национальных валют по отношению к доллару США (в среднем за месяц, квартал, либо год), которые коррелируют во времени с принятием или вступлением
   в силу/отменой сформировавших препятствие нормативных правовых документов одного из государств – членов Союза.

В качестве используемой в анализе информационной базы выступают официальные данные национальной статистики государств-членов и таможенных органов государств-членов, данные статистики ЕАЭС. Длительность анализируемых динамических рядов составляет не менее двадцати периодов, включая период введения препятствия (как правило месяц введения препятствия). Метод статистической оценки влияния факторов на взаимную торговлю на основе регрессионного анализа включает построение динамической многофакторной модели. Метод предполагает применение следующего **алгоритма оценки** влияния значений динамических рядов факторных признаков (переменных) на значения динамических рядов результирующего фактора по взаимной торговле определенного товара, подпадающего под действие препятствия:

1. В офисном приложении MS Excel формируется массив статистических данных по переменным многофакторной модели (2) на основе построения динамических рядов различных факторов. В качестве результирующей переменной в модели выступают стоимостные/физические объемы экспорта/импорта товаров во взаимной торговле между государствами – членами Союза. В качестве факторных переменных выступают экономические показатели, которые могут оказать влияние на стоимостные/физические объемы экспорта/импорта товаров во взаимной торговле между государствами – членами Союза с учетом ввода препятствия.
2. Осуществляется отбор факторных переменных – экономические показатели (месячной, квартальной, либо годовой периодичности), характеризующие средние цены внутреннего рынка государства-члена (хозяйствующего субъекта государства-члена), значения курсов национальных валют государства-члена к доллару США, физические объемы производства, другие абсолютные (относительные) экономические показатели.
3. Значения динамических рядов переменных, включаемых в модель, строятся за период времени, отражающий явление до и во время сформировавшегося препятствия, содержащий не менее чем двадцать уровней.
4. Построение и анализ динамической модели множественной регрессии осуществляется с использованием свободно распространяемого программного пакета для эконометрического моделирования Gretl (Gnu Regression, Econometrics и TIME-series Library), далее пакет Gretl.
5. Для загрузки массива данных в пакет Gretl требуется открыть пакет и импортировать массив данных, указав в программе путь к файлу с данными.
6. Для оценки степени взаимосвязей между всеми факторами в пакете Gretl строится матрица парных коэффициентов корреляции. Расчет парного коэффициента корреляции осуществляется по формуле: $r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y}$ (значение от -1 до +1). По результатам оценки значений этих коэффициентов отмечается высокая степень зависимости между факторами. Значение коэффициента корреляции ±0,75 и выше в матрице говорит о наличии явления мультиколлинеарности (высокой степени зависимости между факторными переменными). В таком случае при сравнении значений коэффициентов корреляции факторных переменных с результирующей переменной отбирается для включения в регрессионную модель факторная переменная с более высоким значением коэффициента.
7. В этом случае одна из двух факторных переменных со значением ±0,75 и выше, которая в наименьшей степени связана с результирующей переменной, отбирается значение коэффициента корреляции, которое показывает наибольшую степень взаимосвязи факторной переменной x с результирующей (зависимой) переменной y.
8. По отобранным из матрицы значениям парных коэффициентов корреляции формируется уравнение множественной регрессии, определяются факторные признаки, которые требуется включить в уравнение множественной регрессии.
9. В пакете Gretl строится уравнение множественной регрессии по отобранным факторам и характеристики для проверки качества данного уравнения.

Проверка статистического качества оцениваемого уравнения регрессии проводится по следующим направлениям:
- проверка статистической значимости коэффициентов уравнения регрессии;
- проверка общего качества уравнения регрессии;
- проверка свойств данных, выполнимость которых предполагалась при оценивании уравнения (проверка выполнимости предпосылок МНК).

9.1. Проверка статистической значимости коэффициентов уравнения регрессии. Выдвигается нулевая гипотеза о том, что совокупности, рассматриваемые с точки зрения одного или нескольких признаков, одинаковы (утверждение об отсутствии различий в значениях или об отсутствии связи в генеральной совокупности). При этом в общем контексте под нулевой гипотезой принимается утверждение, согласно которому не существует связи между двумя измеряемыми явлениями или группами. Предполагается, что действительное значение равно нулю, а найденное из эксперимента отличие от нуля носит случайный характер. Нулевые гипотезы проверяются с помощью статистических критериев Стьюдента и Фишера. Проверка достоверности коэффициентов модели осуществляется с помощью критерия Стьюдента.

Вычисляется фактическое значение критерия Стьюдента по формуле $t = b_j/S_{bj}$ и сравнивается с табличным значением критерия Стьюдента $t_{\propto/2;\nu}$ на соответствующем уровне значимости α и числу степеней свободы ν = n–m–1; $t_{\text{кр}=}t_{\propto/2;\nu}$. Если расчетное значение критерия Стьюдента t превышает его табличное значение, то соответствующий коэффициент регрессии b_j значим, что показывает степень влияния факторной переменной x на результирующую переменную y.

9.2. Проверка общего качества уравнения регрессии. Проверка достоверности модели производится с помощью статистического критерия Фишера по формуле $F = S_{\text{рег}^2}/S_{\text{ост}^2}$ и сравнивается с табличным значением критерия Фишера $F_{\text{кр}}(\alpha = 0{,}05; m_1 = \kappa - 1;\ m_2 = \text{n} - \text{к})$, где α - уровень значимости критерия (величина, используемая для оценки истинности некоторого результата или гипотезы); m_1 — число степеней свободы для большей дисперсии регрессии; m_2 — число степеней свободы для меньшей дисперсии остатков; n – l объем выборки; κ — количество всех коэффициентов модели. Уровень значимости α означает

пороговый уровень статистической значимости, характеризующий вероятность отвергнуть нулевую гипотезу (вероятность совершения ошибки). Чем меньше уровень значимости α тем меньше риск совершения этой ошибки. На практике часто устанавливаются следующие уровни статистической значимости: $\alpha = 0{,}01; 0{,}05; 0{,}1$.

Если F > Fкр, то нулевая гипотеза отвергается и что модель является достоверной,
Если F < Fкр, то нулевая гипотеза принимается и считается, что достоверность модели не доказана.

9.3. Проверка свойств данных, выполнимость которых предполагалась при оценивании уравнения (проверка выполнимости предпосылок МНК). Для принятия решения о значимости коэффициента также предлагается сравнить рассчитанное p-значение

с выбранным уровнем значимости α, где p-значение - это вероятность получения наблюдаемых результатов при условии, что нулевая гипотеза верна. Если p-значение больше α, то нулевую гипотезу следует принять (коэффициент незначимый), если меньше, то нулевую гипотезу следует отклонить (коэффициент значимый). P-значение для проверки нулевой гипотезы состоящей в том, что $\beta = 0$ (коэффициент является незначимым). Для уровня значимости α решение принимается следующим образом: если p-значение ≥ α, то нулевая гипотеза принимается, коэффициент β –

незначим; если p-значение $\leq \alpha$, то нулевая гипотеза отклоняется, коэффициент β – значим.

10. Осуществляется расчет коэффициента детерминации, который определяет долю дисперсии результативного признака, обусловленную изменением факторного признака и дает оценку относительной меры влияния фактора на результат, фиксируя одновременно и роль ошибок в модели. Коэффициент детерминации (скорректированный коэффициент множественной корреляции) для каждого уравнения вычисляется по следующей формуле: $\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)}$; где n – число наблюдений, m – число параметров при переменных по динамическому ряду, $R^2 = 1 - \frac{\sigma^2}{\sigma_y^2}$ – коэффициент детерминации (коэффициент множественной корреляции), где σ^2 – дисперсия изучаемого признака. Коэффициент детерминации измеряется в пределах от 0 до 1 и показывает какая часть изменения зависимой переменной y определяется всеми входящими в модель объясняющими переменными x. Чем выше показатель детерминации, тем лучше модель описывает исходные данные.

Соответственно $1 - R^2$ характеризует влияние прочих неучтенных факторов в модели.

11. Для проверки отсутствия влияния остатков случайных отклонений выполняется тестирование на отсутствие в модели гетероскедастичности и автокорреляции остатков случайных отклонений для подтверждения правильности оценок качества всего уравнения в целом. Наличие автокорреляции в модели устанавливается с помощью статистики Дарбина-Уотсона по формуле $DW = \frac{\sum(e_i - e_{i-1})^2}{\sum e_i^2}$. Наличие автокорреляции в пакете Gretl проверяется с использованием теста Бройша-Годфри. Если между наблюдениями ряда случайных остатков наблюдается корреляция, то в уравнении $e_t = \rho_1 \cdot e_{t-1} + \rho_2 \cdot e_{t-2} + v_t$, где t=1, …, n, коэффициент ρ значимо отличается от 0. Это говорит об отсутствии автокорреляции остатков и как следствие учета влияния неучтенных в модели случайных отклонений. При условии значения коэффициента ρ выше критического уровня значимости ($\alpha = 0{,}05$) в анализируемой эконометрической модели отсутствует автокорреляция остатков случайных отклонений.
12. Для проверки модели на наличие гетероскедастичности остатков применяется тест Уайта, Наличие гетероскедастичности в модели, проверяется с использованием теста Уайта. Строится регрессия:

$$\varepsilon^2 = a + b_1 x_1 + b_{11} x_1^2 + b_2 x_2 + b_{22} x_2^2 + b_{12} \cdot x_1 \cdot x_2$$

Результаты теста Уайта показывают отсутствие гетероскедастичности, в случае если значение ρ-вероятности превышает уровень значимости ($\alpha = 0{,}05$), то можно судить об отсутствии гетероскедастичности остатков, что подтверждает правильность выбора построения эконометрической модели с помощью классического метода МНК.
13. Оценка влияния факторов на результирующую переменную позволяет учитывать значения коэффициентов регрессионной модели. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на зависимую переменную из-за различия единиц измерения и разной степени колеблемости. Для устранения таких различий при интерпретации применяются средние коэффициенты эластичности, определяемые по формуле $Э_j = b_j \cdot \frac{\overline{x_j}}{\overline{y}}$.

Средние коэффициенты эластичности позволяют сравнивать факторы в процентах для всех видов функциональных уравнений. Коэффициенты эластичности показывают, на сколько процентов в среднем изменится зависимая переменная при

изменении факторной переменной на один процент при фиксированных значениях других факторных переменных.

14. Алгоритм применения метода статистической оценки влияния факторов на взаимную торговлю на основе регрессионного анализа завершается построением модели динамического ряда зависимой переменной y – стоимостного/физического объема взаимной торговли от факторных признаков путем ввода фактических значений факторов в получившееся уравнение эконометрической модели. По полученным данным делается оценка роста/снижения результативного признака при фиксации роста/снижения значений факторных переменных по рассчитанным коэффициентам регрессии в модели. Расчетные данные по уравнению эконометрической модели зависимой переменной y сравниваются с фактическими данными (для наглядности строится графическое изображение). Сравнение указанных значений позволяет дать оценку объемов взаимной торговли, недополученных в ходе осуществления внешнеэкономической деятельности между государствами-членами Союза.