

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x) \cdot \text{Var}(y)}}$$

Эконометрическое моделирование

Лабораторная работа № 5

Множественная регрессия



Оглавление

| | |
|--|----------|
| Множественная регрессия | 3 |
| Мультиколлинеарность | 4 |
| Задание 1. Построение модели множественной регрессии | 5 |

Множественная регрессия

Множественная регрессия широко используется в макроэкономических расчетах и является одним из наиболее распространенных методов в эконометрике. Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Модель множественной регрессии описывает, как изменяется в среднем значение зависимой переменной y (результативного признака) при изменении значений объясняющих переменных (факториальных признаков).

По характеру взаимосвязей модели могут быть линейными и нелинейными как по переменным, так и по параметрам.

Для оценки параметров модели методом наименьших квадратов нелинейные по параметрам модели должны быть приведены к линейному виду, часто путем логарифмирования.

Основные предпосылки модели линейной регрессии такие же, как и в парной регрессии. Важным дополнительным условием является некоррелированность объясняющих переменных между собой (отсутствие мультиколлинеарности). Считается, что явление мультиколлинеарности наблюдается тогда, когда коэффициент корреляции между объясняющими переменными превышает по модулю 0,7.

Количество наблюдений при построении модели множественной регрессии должно не менее чем в 3 раза (лучше в 5-6 раз) превышать число оцениваемых параметров.

Уравнение линейной множественной регрессии имеет вид:

$$y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2 + \dots + \hat{\beta}_m \cdot x_m + \varepsilon$$

Коэффициент регрессии $\hat{\beta}_i$ показывает, на сколько единиц изменится величина зависимой переменной y при изменении значения i -й объясняющей переменной на одну единицу при прочих равных условиях (все остальные объясняющие переменные равны своим средним значениям).

Оценка значимости коэффициентов регрессии осуществляется, как и в парной регрессии, по критерию Стьюдента.

Качество построенной модели также оценивает коэффициент детерминации ². При линейной зависимости коэффициент детерминации равен квадрату коэффициента множественной корреляции. Чтобы учесть влияние числа факторов в модели, используют скорректированный коэффициент детерминации (для модели той же формы и числа наблюдений).

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n - 1}{n - m - 1} = R^2 - \frac{m}{n - m - 1} (1 - R^2)$$

m – число факторов, n – число наблюдений.

Значимость уравнения множественной регрессии в целом оценивается с помощью критерия Фишера

$$F_{набл} = \frac{R^2(n - m - 1)}{(1 - R^2)m}$$

С помощью критерия Фишера можно анализировать проблему спецификации модели, в частности, добавление или исключение объясняющей переменной из модели.

Для этого на основе тех же n наблюдений строятся две модели – одна с прежним числом переменных m , а другая – с новым числом $(m+q)$, где q – переменные, включенные в модель.

Для обоснования включения новых q переменных в модель рассчитывается F-статистика

$$F_{\text{набл}} = \frac{R_2^2 - R_1^2}{1 - R_2^2} \cdot \frac{n - m - 1}{q}$$

Где R_2^2 – коэффициент детерминации в модели с большим числом переменных,

R_1^2 – коэффициент переменных в модели с меньшим числом переменных.

Если $F_{\text{набл}}$ превышает $F_{\text{крит}}(q, n-m-1)$, то включение новых переменных целесообразно, они объясняют существенную часть вариации y , не объясненную в модели с меньшим числом переменных.

Существуют две методики построения множественной регрессии – пошаговая вперед и пошаговая назад.

Пошаговая вперед заключается в то, что первоначально строится модель с одной экзогенной переменной. Затем добавляется следующая и строится новая модель. Модели сравниваются и, в зависимости от того ухудшилась или улучшилась модель, введенная переменная либо остается в модели, либо заменяется на другую. Таким образом, перебираются различные комбинации экзогенных переменных, в результате получается наилучшая модель.

Пошаговая назад начинается с того, что рассчитывается множественная регрессия на всем множестве факторов. Затем построенная модель исследуется с точки зрения статистической значимости модели в целом, статистической значимости коэффициентов регрессии, оценивается коэффициент детерминации. Затем из модели удаляется один из влияющих факторов. Его выбор можно осуществить следующим образом:

1. Строится матрица парных коэффициентов корреляции между переменными.
2. Выбираются две экзогенные переменные, между которыми наибольший коэффициент парной корреляции.
3. Из этих двух переменных выбирается та, которая оказывает меньшее влияние на эндогенную переменную, и исключается из модели.

Затем строится новая модель, исследуется ее качество. Также проводится тест на лучшую из двух моделей: с меньшим или большим числом переменных. В конце получается наилучшая модель.

Мультиколлинеарность

Частичная мультиколлинеарность – наличие сильной линейной корреляционной связи между регрессорами.

Основное негативное последствие – стандартные ошибки оценок коэффициентов оказываются высокими. Точность оценивания оказывается низкой. Та или иная степень

корреляции между регрессорами существует всегда. Проблема возникает только когда эта линейная связь проявляется слишком сильно.

Мультиколлинеарность можно выявить на начальном этапе моделирования (до построения регрессии). О ней могут свидетельствовать:

1. Большие (по абсолютной величине) парные коэффициенты корреляции между независимыми переменными.
2. Высокие (>10) значения коэффициента VIF.

Коэффициент VIF (variance inflation factor) характеризует силу мультиколлинеарности. Вычисляется на основе значений R^2 во вспомогательных регрессиях одного регрессора на другие:

$$x_i^k = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \dots + \beta_{k-1} x_i^{(k-1)} + u_i$$

$$VIF = \frac{1}{1 - R^2}$$

Симптомами присутствия мультиколлинеарности в уже построенной модели являются:

1. Небольшое изменение исходных данных, приводит к существенному изменению оценок коэффициентов.
2. Каждая переменная в отдельности является незначимой, а уравнение в целом имеет высокий R^2 и является значимым.
3. Оценки коэффициентов имеют неправильные с точки зрения экономической теории знаки или неоправданно большие значения.

Задание 1. Построение модели множественной регрессии

Имеются данные о макроэкономических показателях России за 1996-2013 гг. (данные представлены в таблице 1).

y – среднемесячная заработная плата (руб.)

x_1 – ВВП (млрд. руб.)

x_2 – объем отгруженной продукции, выполненных работ и оказанных услуг (млрд. руб.)

x_3 – экспорт (млрд. долларов США)

x_4 – бюджетный дефицит (млрд. руб.)

x_5 – курс доллара (руб. за доллар)

x_6 – индекс цен (инфляция)

Данные находятся в [файле Excel](#).

Требуется определить факторы, влияющие на среднемесячную заработную плату.