Федеральное агентство по образованию и науке РФ

Костромской государственный технологический университет.

Кафедра высшей математики

Реферат

по эконометрике на тему:

Мультиколлинеарность

Выполнила

студент 1 курса

заочного факультета

сп-ть «Бухгалтерский учёт,

анализ и аудит».

Проверила

Катержина С.Ф.

Кострома 2008 г

**Мультиколлинеарность**

Под мультиколлинеарностью понимается высокая взаимная коррелированность объясняющих переменных. Мультиколлинеарность может проявляться в функциональной (явной) и стохастической (скрытой) формах.

При функциональной форме мультиколлинеарности по крайней мере одна из парных связей между объясняющими переменными является линейной функциональной зависимостью. В этом случае матрица X`X особенная, так как содержит линейно зависимые векторы-столбцы, и её определитель равен нулю, т.е. нарушается предпосылка регрессионного анализа, это приводит к невозможности решения соответствующей системы нормальных уравнений и получения оценок параметров регрессионной модели.

Однако в экономических исследованиях мультиколлинеарность чаще проявляется в стохастической форме, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Матрица X`X в этом случае является неособенной, но её определитель очень мал.

В то же время вектор оценок b и его ковариционная матрица ∑b пропорциональны обратной матрице (X`X)-1, а значит, их элементы обратно пропорциональны величине определителя |X`X|. В результате получаются значительные средние квадратические отклонения (стандартные ошибки) коэффициентов регрессии b0, b1,…,bp и оценка их значимости по t-критерию не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F-критерию.

Оценки становятся очень чувствительными к незначительному изменению результатов наблюдений и объёма выборки. Уравнения регрессии в этом случае, как правило, не имеют реального смысла, так как некоторые из его коэффициентов могут иметь неправильные с точки зрения экономической теории знаки и неоправданно большие значения.

Точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее, имеются некоторые эвристические подходы по её выявлению.

Один из таких подходов заключается в анализе корреляционной матрицы между объясняющими переменными X1,X2,…,Xp и выявлении пар переменных, имеющих высокие переменные корреляции (обычно больше 0,8). Если такие переменные существуют, говорят о мультиколлинеарности между ними. Полезно также находить множественные коэффициенты детерминации между одной из объясняющих переменных и некоторой группой из них. Наличие высокого множественного коэффициента детерминации (обычно больше 0,6) свидетельствует о мультиколлинеарности.

Другой подход состоит в исследовании матрицы X`X. Если определитель матрицы X`X либо её минимальное собственное значение λmin близки к нулю ( например одного порядка с накапливающимися ошибками вычислений), то это говорит о наличии мультиколлинеарности. о том же может свидетельствовать и значительное отклонение максимального собственного значения λmax матрицы X`X от её минимального собственного значения λmin.

Для устранения или уменьшения мультиколлинеарности используется ряд методов. Самый простой из них (но далеко не всегда возможный) состоит в том, что из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. При этом, какую переменную оставить, а какую удалить из анализа, решают в первую очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

Другой метод устранения или уменьшения мультиколлинеарности заключается в переходе от несмещённых оценок, определённых по методу наименьших квадратов, к смещённым оценкам, обладающим, однако, меньшим рассеянием относительно оцениваемого параметра, т.е. меньшим математическим ожиданием квадрата отклонения оценки bj от параметра βj или M (bj - βj)2.

Оценки, определяемые вектором, обладают в соответствии с теоремой Гаусса-Маркова минимальными дисперсиями в классе всех линейных несмещённых оценок, но при наличии мультиколлинеарности эти дисперсии могут оказаться слишком большими, и обращение к соответствующим смещённым оценкам может повысить точность оценивания параметров регрессии. На рисунке показан случай, когда смещённая оценка βj^, выборочное распределение которой задаётся плотностью φ ( βj^).

Действительно, пусть максимально допустимый по величине доверительный интервал для оцениваемого параметра βj есть (βj-Δ, βj+Δ). Тогда доверительная вероятность, или надёжность оценки, определяемая площадью под кривой распределения на интервале (βj-Δ, βj+Δ), как нетрудно видеть из рисунка, будет в данном случае больше для оценки βj по сравнению с bj ( на рисунке эти площади заштрихованы). Соответственно средний квадрат отклонения оценки от оцениваемого параметра будет меньше для смещённой оценки, т.е.:

M ( βj^- βj )2 < M ( bj - βj )2

При использовании «ридж-регрессии» (или «гребневой регрессии») вместо несмещённых оценок рассматривают смещённые оценки, задаваемые вектором

βτ^=( X`X+τ Ep+1)-1 X`Y,

где **τ –** некоторое положительное число, называемое «гребнем» или «хребтом»,

Ep+1 – единичная матрица (р+1) –го порядка.

Добавление **τ** к диагональным элементам матрицы X`X делает оценки параметров модели смещёнными, но при этом увеличивается определитель матрицы системы нормальных уравнений – вместо (X`X) от будет равен

|X`X+τ Ep+1|

Таким образом, становится возможным исключение мультиколлинеарности в случае, когда определитель |X`X| близок к нулю.

Для устранения мультиколлинеарности может быть использован переход от исходных объясняющих переменных X1,X2,…,Xn, связанных между собой достаточно тесной корреляционной зависимостью, к новым переменным, представляющим линейные комбинации исходных. При этом новые переменные должны быть слабо коррелированными либо вообще некоррелированными. В качестве таких переменных берут, например, так называемые главные компоненты вектора исходных объясняющих переменных, изучаемые в компонентном анализе, и рассматривают регрессию на главных компонентах, в которой последние выступают в качестве обобщённых объясняющих переменных, подлежащих в дальнейшем содержательной (экономической) интерпритации.

Ортогональность главных компонент предотвращает проявление эффекта мультиколлинеарности. Кроме того, применяемый метод позволяет ограничиться малым числом главных компонент при сранительно большом количестве исходных объясняющих переменных.

*Мультиколлинеарность —* это понятие, которое используется для описания проблемы, когда нестрогая линейная зависимость между объясняющими переменными приводит к получению ненадежных оценок регрессии. Разумеется, такая зависимость совсем необязательно дает неудовлетворительные оценки. Если все другие условия благоприятствуют, т. е. если число наблюдений и выборочные дисперсии объясняющих переменных велики, а дисперсия случайного члена —мала, то в итоге можно получить вполне хорошие оценки.

Итак, мультиколлинеарность должна вызываться сочетанием нестрогой зависимости и одного (или более) неблагоприятного условия, и это — вопрос

степени выраженности явления, а не его вида. Оценка любой регрессии будет страдать от нее в определенной степени, если только все независимые переменные не окажутся абсолютно некоррелированными. Рассмотрение данной проблемы начинается только тогда, когда это серьезно влияет на результаты оценки регрессии.

Эта проблема является обычной для регрессий временных рядов, т. е. когда данные состоят из ряда наблюдений в течение какого-то периода времени. Если две или более независимые переменные имеют ярко выраженный временной тренд, то они будут тесно коррелированы, и это может привести к мультиколлинеарности.

***Что можно предпринять в этом случае?***

Различные методы, которые могут быть использованы для смягчения мультиколлинеарности, делятся на две категории: к первой категории относятся попытки повысить степень выполнения четырех условий, обеспечивающих надежность оценок регрессии; ко второй категории относится использование внешней информации. Если сначала использовать возможные непосредственно получаемые данные, то, очевидно, было бы полезным увеличить число наблюдений.

Если вы применяете данные временных рядов, то это можно сделать путем сокращения продолжительности каждого периода времени. Например, при оценивании уравнений функции спроса в упражнениях 5.3 и 5.6 можно перейти с использования ежегодных данных на поквартальные данные.

После этого вместо 25 наблюдений их станет 100. Это настолько очевидно и так просто сделать, что большинство исследователей, использующих временные ряды, почти автоматически применяют поквартальные данные, если они имеются, вместо ежегодных данных, даже если проблема мультиколлинеарности не стоит, просто для сведения к минимуму теоретических дисперсий коэффициентов регрессии. В таком подходе существуют, однако, и потенциальные проблемы. Можно привнести или усилить автокорреляцию, но она может быть нейтрализована. Кроме того, можно привнести (или усилить) смещение, вызванное ошибками измерения, если поквартальные данные измерены с меньшей точностью, чем соответствующие ежегодные данные. Эту проблему не так просто решить, но она может оказаться несущественной.

Если вы используете данные перекрестной выборки и находитесь на стадии планирования исследования, то можно увеличить точность оценок регрессии и ослабить проблему мультиколлинеарности просто за счет большего расхода средств на увеличение размера выборки. Однако такой подход имеет уменьшающуюся предельную отдачу, поскольку стандартные отклонения коэффициентов регрессии обратно пропорциональны величине 7л > в то время как расходы прямо пропорциональны *п.*

Столь же важно, если вы используете данные перекрестной выборки и находитесь на стадии планирования исследования, максимизировать дисперсию наблюдений независимых переменных в выборке, например путем расслоения выборки.

Далее, можно сократить величину а^. Случайный член включает в себя объединенный эффект всех переменных, оказывающих влияние на величину *уу* которые не включены явно в уравнение регрессии. Если вы допускаете мысль о том, что важная переменная могла быть опущена и, следовательно, оказывает влияние на и, то можно сократить величину а^, если добавить эту переменную в уравнение регрессии.

Если, однако, новая переменная линейно связана с одной или несколькими переменными, уже включенными в уравнение, то ее введение может еще больше усугубить проблему мультиколлинеарности. Мы вернемся к обсуждению этого вопроса, который представляет большую практическую важность, в конце следующей главы после рассмотрения ошибок спецификации.

Наконец, об использовании самого простого метода. Если вы действительно имеете возможность собрать дополнительные данные, то нужно постараться получить выборку, в которой независимые переменные слабо связаны между собой.

Существуют два типа внешней информации, которая может оказаться полезной: теоретические ограничения и внешние эмпирические оценки. Теоретическое ограничение представляет собой допущение, касающееся величины коэффициента или некоторой связи между коэффициентами. Поясним это на примере.

При построении производственной функции с использованием данных временных рядов (как это было сделано в разделе 5.3) следует иметь в виду, что на выпуск продукции, наряду с изменениями в капитальных и трудовых затратах, вероятно, будет оказывать влияние технический прогресс. Если вы имеете дело с агрегированными данными, то невозможно количественно оценить технический прогресс, и проще всего включить экспоненциальный временной тренд в уравнение, записав функцию Кобба—Дугласа, например, в виде:

***Г= AK«Lte«v,* (5.47)**

где *Y, К* и *L* имеют те же определения, что и в разделе 5.3; / — время; *г* — темп

прироста выпуска благодаря техническому прогрессу. Оценив это соотношение по данным табл. 5.1, получим (стандартные ошибки указаны в скобках):

log У= 2,81 - 0,53 log *K+* 0,91 log *L* + 0,047/; *R2* = 0,97; (5.48)

(1,38) (0,34) (0,14) (0,021) *F=* 189,8.

Со всей очевидностью этот результат показывает, что эластичность выпуска продукции по затратам капитала отрицательна, что означает снижение выпуска при увеличении затрат капитала. Уравнение также показывает темп прироста выпуска продукции за счет технического прогресса порядка 4,7% в год, что является неправдоподобно высокой оценкой для рассматриваемого периода.

Здесь можно предположить, что по крайней мере отчасти проблема связана с мультиколлинеарностью, так как коэффициент корреляции между log *К и t* составляет 0,997, а стандартная ошибка коэффициента при log *Къ* 5 раз больше, чем в уравнении без величины / (5.32).

Отсюда появляется желание ввести ограничения на эффект от масштаба, рассматривая его как постоянную величину, что позволит переписать уравнение только с двумя независимыми переменными, имеющими временной тренд, вместо трех и с капиталовооруженностью труда в качестве объясняющей переменной вместо затрат капитала. Этот показатель по-прежнему тесно коррелирован с временем (коэффициент корреляции составляет 0,96), но степень коррелированности уже не так предельно высока. Оценив уравнение (5.28) с экспоненциальным временным трендом, мы получим (стандартные ошибки указаны в скобках):

log *Y/L =* -0,11 + 0,11 log *K/L* + 0,006f; *R2* = 0,65; (5.49)

(0,03) (0,15) (0,006) F=19,5.

Оценки величин а и г, хотя и незначимо отличаются от нуля, теперь более реалистичны, чем раньше, а стандартные ошибки — намного меньше, чем в уравнении (5.48). Тот факт, что величина г незначимо отличается от нуля, подтверждает вывод Ч. Кобба и П. Дугласа о том, что темп увеличения общей производительности факторов в рассматриваемый период был очень низким. Очевидно, что обоснованность этой процедуры зависит от правильности введенного ограничения, поэтому сначала нужно статистически проверить ограничение, что рассматривается в следующей главе.

Наконец, можно использовать внешние оценки. Предположим, что вы решили воспользоваться уравнением (5.24) в качестве формулы для функции спроса, но имеется проблема мультиколлинеарности, так как располагаемый личный доход и цена имеют ярко выраженные временные тренды, а следовательно, тесно коррелированы. Предположим, однако, что вы также имеете перекрестные статистические данные для *у* и х, полученные из другой выборки. Если допустить, что все домохозяйства в проводимом анализе платили за данный товар одинаковую цену, то модель примет вид:

log/=loga' + p\*logx' + w'. (5.50)

Получив оценку *Ь\* для р\ при оценивании регрессионной зависимости *у* от *х\* вы можете подставить ее в уравнение (5.24). Теперь определяется новая переменная logy, равная (log *у —* 6',log х), описывающая спрос, скорректированный на изменения дохода. После этого уравнение (5.25) принимает вид:

logy = log a+ p2 log/J +w. (5.51)

Рассчитав logy для каждого наблюдения, вы оцениваете его регрессионную зависимость от log *p,* и, так как здесь имеется только одна независимая переменная, мультиколлинеарность автоматически исключается.

При использовании этого метода могут возникнуть две проблемы, которые необходимо учитывать. Во-первых, оценка величины Р2 зависит от точности оценки величины Р'р которая, безусловно, подвержена влиянию ошибки выборки.

Во-вторых, вы допускаете, что коэффициент при доходе имеет одинаковый смысл для случаев временных рядов и перекрестных выборок, что, конечно, может быть и не так. Для большинства товаров краткосрочная и долгосрочная эластичность спроса по доходу может значительно различаться. Одна из причин этого состоит в том, что характер расходов подвержен влиянию инерции, которое в краткосрочном периоде может превзойти эффекты дохода.

Другая причина заключается в том, что изменение уровня дохода может оказать на расходы как непосредственное (в виде изменения бюджетного ограничения), так и косвенное влияние (за счет изменения образа жизни), причем косвенное влияние происходит намного медленнее, чем прямое. В качестве первого приближения обычно считается, что регрессии для временных рядов, особенно с небольшими периодами выборки, дают показатели краткосрочной эластичности, в то время как регрессии с использованием данных перекрестных выборок дают показатели долгосрочной эластичности.

**Список используемой литературы:**

1. Н.Ш. Кремер, Б.А. Путко «Эконометрика», Москва, 2005

2. К. Доугерти «Введение в эконометрику», Москва, 1999