

Регрессии с эндогенностью: инструментальные переменные

**По мотивам статьи
П.Эббеса
«Инструментальные
переменные:
нетехнический анализ»**

Проблема

Важным предположением КЛРМ является предположение

экзогенности регрессоров,

то есть некоррелированности регрессоров и случайной ошибки.

Если в регрессионной модели $Y = Xb + e$ регрессоры X коррелируют с ошибкой e , они называются **эндогенными**

Проблема

- Обычные оценки метода наименьших квадратов в этом случае
 - смещены и
 - несостоятельны
- Содержательная интерпретация ошибочна
- Рекомендации, выработанные на основе модели неверны

Причины эндогенности

1. Пропуск существенных переменных
2. Ошибки измерения регрессоров
3. Самоотбор
4. Одновременность
5. Автокорреляция ошибок при наличии в уравнении лага зависимой переменной в роли регрессора

Причины эндогенности

- Причины 2-5 по сути есть следствия причины 1
- Разные источники эндогенности могут иметь место одновременно
- Источники могут как усиливать, так и компенсировать друг друга

Пропуск существенной переменной

Анализ рынка труда

- Пропуск переменной «способности» при оценивании эффекта образования на заработную плату:
- «способности» сильно и положительно коррелируют с образованием и, будучи пропущенными, приводят к смещению оценки эффекта образования вверх

Пропуск существенной переменной

Маркетинговые исследования

При анализе влияния **цен** или **рекламы** на **объемы продаж** пропускаются важные, но ненаблюдаемые характеристики локальных рынков или товаров:

- уровень конкуренции,
- слухи,
- изменения во вкусах,
- доли различных игроков,
- наличие скидок,

которые, будучи коррелированными с ценами и реакцией потребителя на рекламу, приводят к смещению оценок влияния последних

Ошибки измерения

Причины ошибок измерения

- Метод измерения
- Инструмент измерения
- Отсутствие физической единицы измерения
- Ошибки агрегирования

Ошибки измерения

- Ошибки измерения образования:
 - число лет обучения не отражает самообразования
- Ошибки измерения переменных в обследованиях домохозяйств
- Ошибки измерения цены в моделях совокупного спроса
 - потребители сталкиваются с разными ценами и товарными характеристиками

Ошибки измерения

- Неудачные шкалы рейтингов для измерения восприятия, вер, отношений, суждений
- Ошибки метода измерения:
 - эффект ореола
 - эффекты интервьюирующего
 - искажения от эффекта социальной желательности

Самоотбор

Возникает, когда индивиды выбирают себе определенное состояние, руководствуясь скрытыми причинами:

- выходить на рынок труда или нет,
- быть или не быть членом профсоюза,
- работать в бюджетном секторе или частном;

или когда опрос ведется через интернет

Одновременность

Проявляется, когда переменные правой части регрессионного уравнения определяются одновременно с зависимыми переменными, т.е. эндогенно порождаются процессом, а не задаются извне (экзогенно)

- принятие решения об образовании или о выходе на рынок труда,
- формирование цены и величины спроса

Авторегрессии с автокоррелированными ошибками

- Эффект продажи каталога на вероятность покупки товара из каталога:
 - переменная почтовой рассылки и другие рекламные воздействия на потребителя – функции прошлых продаж

Методы оценивания

- Метод инструментальных переменных
 - инструментов столько же, сколько регрессоров, для экзогенных регрессоров – инструменты они же сами
- Двухшаговый МНК
 - инструментов больше, чем регрессоров

Инструменты

Некие переменные Z ,
удовлетворяющие требованиям:

- сильной коррелированности с инструментруемыми эндогенными X
- некоррелированности с ошибкой

Проблема выбора инструментов

- Инструменты должны браться вне модели
- Требования к инструментам противоречивы:
- Если Z должны сильно коррелировать с X , а X должны сильно коррелировать с Y , то Z будут сильно коррелировать с Y , а, следовательно, коррелировать с ошибкой

Следствия противоречивости требований к инструментам

- Либо инструменты экзогенные (нет корреляции с ошибкой), но слабые (слабо коррелируют с X)
- Либо инструменты сильные, но эндогенные (сильно коррелируют с ошибкой)

Примеры инструментов

- Для образования
 - образование в семье, месяц рождения, близость колледжа
- Для цены
 - лаги рекламы, лаги цены, данные по ценам других рынков
- Разные инструменты часто приводят к разным результатам

Слабые инструменты

Инструменты являются слабыми,

- Когда они не обладают высокой объясняющей силой по отношению к эндогенным регрессорам
- Когда число инструментов слишком велико

Проблемы использования инструментов

Три «ловушки» слабых инструментов:

- смещение оценок двухшагового МНК в конечных выборках
- эндогенность
- плохое асимптотическое приближение фактического распределения инструментальной оценки

Смещение оценок двухшагового МНК в конечных выборках

- Смещение оценок IV и TSLS оказывается в том же направлении, что и у оценок МНК
- Смещение на конечных выборках является функцией числа инструментов
- Чем больше слабых инструментов, тем больше смещение

Эндогенность

Инструментальная оценка может
оказаться более несостоятельной,
чем оценка МНК,
если инструменты коррелируют с
ошибкой

Неадекватность асимптотических приближений

При слабых инструментах

- асимптотическая приближение ковариационной матрицы оценок будет больше реальной, т.е. оценка обладает большой дисперсией (неэффективна), следовательно
- сложно делать адекватные выводы о статистической значимости коэффициентов

Проверка качества инструментов

- Stock, Wright & Yogo (2002) утверждают следующее: F-статистика первого шага TSLS (для регрессии X на Z) должна быть больше 10,
- тогда можно рассчитывать на достоверность выводов о статистической значимости оценок второго шага TSLS

Проверка качества инструментов

- Широко используемый тест Хаусмана одновременно проверяет экзогенность регрессоров и релевантность инструментов,
- поэтому, когда основная гипотеза отвергается, неясно, что именно стало причиной.

Проверка качества инструментов

- В статье рассматриваются два более удачных, судя по всему, теста:
- Hahn & Hausman (2002)
- Ebbes, Wedel ... (2005)

(последний основан на методе латентных переменных и может быть использован для проверки инструментов на слабость и эндогенность как вместе, так и отдельно)

Выбор числа инструментов

- Лучше не более трех-четырех (для одного эндогенного регрессора)
- Можно использовать метод главных компонент для инструментов и выбирать компоненты, наиболее связанные с X , отбрасывая остальные
- Чем слабее инструменты, тем меньше их должно быть

Решение проблемы

- Если есть панель, то часто инструменты не нужны
- Инструментом может быть переменная группировки (группировка должна различать ненаблюдаемые высокие и низкие значения пропущенной переменной)
- Инструментом может быть лаг высокого порядка
- В СОУ можно использовать легкодоступные экзогенные переменные, не включенные в модель

Экономные методы инструментальных переменных

- Ebbes, Wedel & Böckenholt (2006) предложили 3 метода, не основанные на наблюдаемых инструментах:
- Метод моментов высокого порядка
- Метод идентификации через гетероскедастичность (IV=группировка по гетероскедастичности))
- Метод латентных инструментальных переменных (ММП)