

Эконометрика. Лекция 10

Эндогенность и инструментальные переменные (2)

Д. С. Терещенко

НИУ ВШЭ, Санкт-Петербург

25 ноября 2022 г.

Основные источники

- **SW, глава 12**
- Видео-материалы для самостоятельного освоения

Содержание

- 1 Повторение
- 2 «Множественный-множественный» случай 2МНК
- 3 Инструментальные переменные на практике
- 4 Алгебра 2МНК

Содержание

① Повторение

② «Множественный-множественный» случай 2МНК

③ Инструментальные переменные на практике

④ Алгебра 2МНК

Еще раз об эндогенности

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- Эндогенность: мы не уверены, что $\mathbb{E}(u_i|x_i) = 0$
- Идея борьбы с эндогенностью:
 - 1 эндогенная переменная X раскладывается на две составляющие:
 - проблемная компонента, которая может быть коррелирована с ошибкой,
 - хорошая, очищенная компонента, которая не коррелирована с ошибкой;
 - 2 очищенная компонента используется для получения оценок.

Еще раз об инструментальных переменных и 2МНК

- Для хороших инструментов должно выполняться:
 - релевантность сила: $\text{cov}(z, x) \neq 0$
 - экзогенность: $\mathbb{E}(u_i | z_i) = 0 \Rightarrow \text{cov}(z, u) = 0$
- 2МНК в «парном-парном» случае:
 - 1 $x_i = \pi_0 + \pi_1 z_i + v_i$
 - 2 $y_i = \beta_0 + \beta_1 x_i + u_i$
- Оценки 2МНК:
 - несмещенные
 - состоятельные
 - асимптотически нормальные

$$\hat{\beta}_1^{TSLS} = \frac{\widehat{\text{cov}}(z, y)}{\widehat{\text{cov}}(z, x)} \xrightarrow[n \rightarrow \infty]{p} \frac{\text{cov}(z, y)}{\text{cov}(z, x)} = \frac{\text{cov}(z, \beta_0 + \beta_1 x + u)}{\text{cov}(z, x)} = \beta_1$$
$$\sqrt{n}(\hat{\beta}_1^{TSLS} - \beta_1) \xrightarrow[n \rightarrow \infty]{d} N(0, \dots)$$

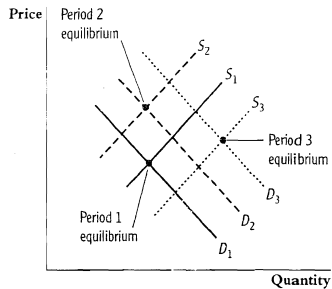
Еще раз о поиске инструментов

При выборе инструментов необходимо полагаться на:

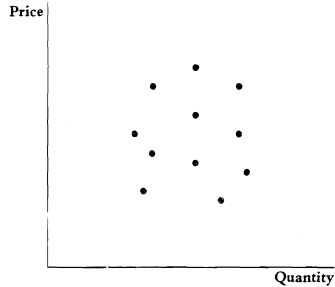
- экономическую теорию;
- экспертные мнения, институциональное знание, экономическую интуицию.

Пример использования экономической теории для поиска инструмента

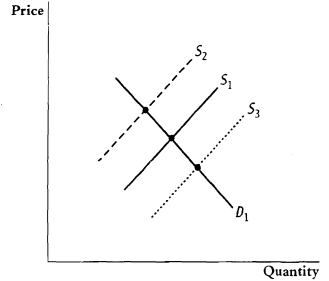
Оценка кривой спроса (Stock J. & Watson M. Introductory econometrics, ch. 12)



(a) Demand and supply in three time periods



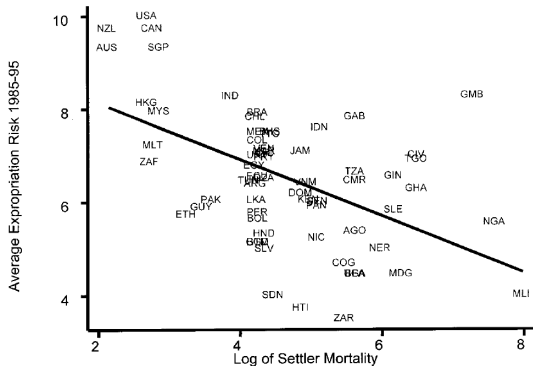
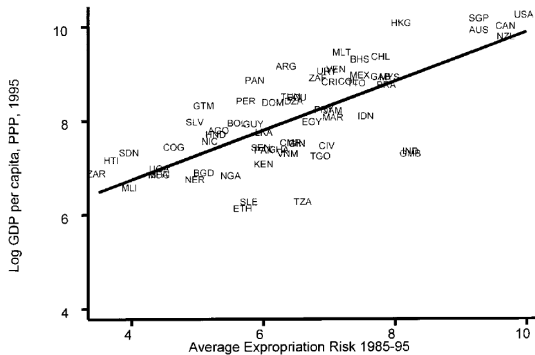
(b) Equilibrium price and quantity for 11 time periods



(c) Equilibrium price and quantity when only the supply curve shifts

Пример использования исторического анализа для поиска инструмента

Институты и экономическое развитие: Acemoglu, Johnson & Robinson (2001)



(potential) settler mortality \Rightarrow settlements \Rightarrow early institutions \Rightarrow current institutions \Rightarrow current performance.

Пример использования исторического анализа для поиска инструмента

Телевидение и результаты выборов: Enikolopov R., Petrova M., Zhuravskaya E. (2011)

- Переменные:
 - Зависимая переменная: решение индивида голосовать или не голосовать за ту или иную партию.
 - Переменная интереса: смотрел ли индивид оппозиционный телеканал.
 - **Эндогенность! Почему?**
- Выборы в Государственную думу РФ (1999):
 - За два месяца до выборов лидером была оппозиционная партия ОВР, в то время как прооправительственная партия «Единство», судя по опросам, имела рейтинг около 0.
 - Результаты выборов радикально отличаются от опросов: «Единство» набирает 23% голосов, а ОВР — 13%.
- Влияние телевидения:
 - Партия «Единство» была поддержана основными телеканалами ОРТ и РТР.
 - Оппозиционная партия ОВР была поддержана телекомпанией НТВ.
 - Телеканалы ОРТ и РТР были доступны на всей территории России. Сигнал НТВ покрывал только часть территории страны — **хороший источник экзогенности!**

Пример использования знания контекста для поиска инструмента

Продлевает ли жизнь агрессивное лечение сердечных приступов? SW, глава 12

- Агрессивное лечение при сердечных приступах обладает большим потенциалом для спасения жизней.
- Можно оценить регрессию продолжительности выживания пациента в зависимости от бинарной переменной, отражающей факт наличия подобного лечения.
- **Эндогенность!**
 - МНК-оценки подвержены смещению: катетеризация сердца «просто так» не назначается пациентам. Данная процедура проводится, если врач и пациент совместно решают, что это будет эффективным.
 - Это решение основано частично на ненаблюдаемых факторах, связанных с характеристиками здоровья, т. е. может коррелировать с ошибкой.
- В работе Макклеллана, Макнейла и Ньюхауса (McClellan, McNeil, and Newhouse, 1994) было предложено рассматривать географические факторы.
- Макклеллан, Макнейл и Ньюхаус использовали в качестве инструментальной переменной разность между расстоянием от дома пациента с сердечным приступом до ближайшей больницы, в которой осуществляется катетеризация сердца, и расстоянием до ближайшей больницы иного типа.

Содержание

- 1 Повторение
- 2 «Множественный-множественный» случай 2МНК
- 3 Инструментальные переменные на практике
- 4 Алгебра 2МНК

Повторение. Формы записи в ММЛР

- Скалярная: $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$
- Векторная: $y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$
- Матричная: $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$

Обобщенная модель с инструментальными переменными

Переменные:

- y — зависимая переменная
- x — проблемные эндогенные регрессоры, k штук
- w — дополнительные регрессоры (экзогенные переменные, контрольные переменные) r штук
- z — инструментальные переменные, m штук

Формы записи «уравнения интереса»:

- Скалярная: $y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \beta_{k+1} w_{1i} + \dots + \beta_{k+r} w_{ri} + u_i$
- Векторная: $y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$
- Матричная: $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$

Соотношение числа инструментов и числа (эндогенных) объясняющих переменных

- $m = k$ — коэффициенты *точно определены* (точная идентификация)
- $m > k$ — коэффициенты *переопределены* (сверхидентификация)
- $m < k$ — коэффициенты *недоопределены* (недоидентификация)

2МНК для «множественного-множественного» случая

Для случая с одной эндогенной объясняющей переменной:

$$\textcircled{1} X_{1i} = \pi_0 + \pi_1 Z_{1i} + \dots + \pi_m Z_{mi} + \pi_{m+1} W_{1i} + \dots + \pi_{m+r} W_{ri} + v_i$$

$$\textcircled{2} Y_i = \beta_0 + \beta_1 \hat{X}_{1i} + \beta_2 W_{1i} + \dots + \beta_{1+r} W_{ri} + u_i$$

Для случая с несколькими эндогенными объясняющими переменными:

- ① Процедура первого шага повторяется для каждой эндогенной переменной
- ② Регрессия Y_i на оценки эндогенных объясняющих переменных, полученные на первом шаге $(\hat{X}_{1i} \dots \hat{X}_{ki})$, + все W

Предположения модели с инструментальными переменными (простыми словами)

- W должны быть либо экзогенными ($E(u_i|W_i) = 0$), либо контрольными ($E(u_i|Z_i, W_i) = E(u_i|W_i)$)
- $(X_{1i}, \dots, X_{ki}, W_{1i}, \dots, W_{ri}, Z_{1i}, \dots, Z_{mi}, Y_i) \sim i.i.d.$
- Большие выбросы маловероятны: переменные X, W, Z, Y имеют ненулевые конечные четвертые моменты
- Должны выполняться два условия допустимости инструментов (сила и годность)

«Требования» к инструментам в общем случае

- релевантность сила: $\mathbb{E}[\mathbf{z}_i \mathbf{x}_i']$ — матрица полного ранга
- экзогенность goodness: $\mathbb{E}[u_i | \mathbf{z}_i] = 0$
- избыточность: $\mathbb{E}[\mathbf{z}_i \mathbf{z}_i'] \equiv \mathbf{Q}_{ZZ}$ — матрица полного ранга

Содержание

- 1 Повторение
- 2 «Множественный-множественный» случай 2МНК
- 3 Инструментальные переменные на практике**
- 4 Алгебра 2МНК

Проблемы, связанные с выбором правильных инструментов

Cameron & Trivedi (2005). Microeconometrics

IV is widely used in econometrics and rarely used elsewhere, is conceptually difficult and easily misused.

SW, стр. 458–459

Степень полезности регрессии с инструментальными переменными в рамках того или иного приложения зависит от того, являются ли инструменты допустимыми. Неправильный выбор инструментов приводит к бессмысленным результатам. Поэтому важно оценить, является ли данный набор инструментов допустимым в данной конкретной ситуации.

Проблема слабых инструментов

- Инструменты, которые объясняют малую часть изменений X , называют слабыми инструментами.
- Если инструменты являются слабыми, то нормальное распределение является плохим приближением выборочного распределения 2МНК-оценки, даже если размер выборки достаточно велик.
- Важна теория, но для иллюстрации можно проверять гипотезы о значимости инструментов на первом шаге.
- На практике возникает вопрос: насколько сильными должны быть инструменты, чтобы обеспечить хорошее их приближение нормальным распределением на практике?
- Варианты решения проблемы:
 - В случае небольшого числа сильных инструментов и большого числа слабых инструментов лучше отбросить слабые инструменты и использовать наиболее подходящие сильные инструменты для анализа с помощью 2МНК.
 - При точной идентификации или недостаточном количестве сильных инструментов: либо искать новые сильные инструменты, либо использовать более продвинутые по сравнению с 2МНК методы.

Проблема не годных инструментов

- Если инструменты не являются годными, то 2МНК-оценки не являются состоятельными.
- Важна теория, экспертные мнения и т. п.
- При точной идентификации никакой эмпирическая проверка невозможна.
- При сверхидентификации иногда проводят тест на сверхидентифицирующие ограничения (в нашем курсе мы его не проходим).

Содержание

- 1 Повторение
- 2 «Множественный-множественный» случай 2МНК
- 3 Инструментальные переменные на практике
- 4 Алгебра 2МНК

Повторение. Предположения ММЛР

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

- ① $\mathbb{E}(u_i | \mathbf{x}_i) = 0 \quad (i = 1, 2, \dots, n)$
- ② $\text{var}(u_i | \mathbf{x}_i = \mathbf{x}) = \sigma^2(\mathbf{x}) < \infty, \forall \mathbf{x}$
- ③ $\{(y_i, \mathbf{x}_i, u_i)\} \sim i.i.d.$
- ④ $\mathbb{E}[\mathbf{x}_i \mathbf{x}_i'] \equiv \mathbf{Q}_{\mathbf{X}\mathbf{X}}$ — матрица полного ранга

Повторение. МНК

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\mathbf{Q}}_{\mathbf{X}\mathbf{X}}^{-1}\hat{\mathbf{Q}}_{\mathbf{X}\mathbf{y}}$$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow[n \rightarrow \infty]{d} \mathbb{N}(\mathbf{0}_{k+1}, \mathbf{V}_{\hat{\boldsymbol{\beta}}})$$

$$\mathbf{V}_{\hat{\boldsymbol{\beta}}} = n^{-1}\mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1} \mathbb{E} [u_i^2 \mathbf{x}_i \mathbf{x}_i'] \mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}$$

Двухшаговый МНК

Общая схема

$$y = X\beta + u$$

① $X = Z\Pi + v$

② $y = \hat{X}\beta + u$

- $\hat{\Pi} = (Z'Z)^{-1}Z'X = \hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}$

- $\hat{X} = Z\hat{\Pi}$

- $\hat{\beta}^{TSLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$

Двухшаговый МНК

Основные матрицы

$$\mathbf{Z}_{n(1+m+r)} = \begin{bmatrix} 1 & z_{11} & \dots & z_{m1} & w_{11} & \dots & w_{r1} \\ 1 & z_{12} & \dots & z_{m2} & w_{12} & \dots & w_{r2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & z_{1n} & \dots & z_{mn} & w_{1n} & \dots & w_{rn} \end{bmatrix}$$

$$\hat{\mathbf{X}}_{n(1+k+r)} = \begin{bmatrix} 1 & \hat{x}_{11} & \dots & \hat{x}_{k1} & w_{11} & \dots & w_{r1} \\ 1 & \hat{x}_{12} & \dots & \hat{x}_{k2} & w_{12} & \dots & w_{r2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & \hat{x}_{1n} & \dots & \hat{x}_{kn} & w_{1n} & \dots & w_{rn} \end{bmatrix}$$

$$\hat{\mathbf{P}}_{(1+m+r)(1+k+r)} = \begin{bmatrix} 1 & \hat{\pi}_0^1 & \dots & \hat{\pi}_0^k & 1 & \dots & 0 \\ 0 & \hat{\pi}_1^1 & \dots & \hat{\pi}_1^k & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \hat{\pi}_{m+r}^1 & \dots & \hat{\pi}_{m+r}^k & 0 & \dots & 1 \end{bmatrix}$$

Двухшаговый МНК

Формула оценок

$$y = X\beta + u$$

① $X = Z\Pi + V$

② $y = \hat{X}\beta + u$

- $\hat{\Pi} = (Z'Z)^{-1}Z'X = \hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}$

- $\hat{X} = Z\hat{\Pi}$

- $\hat{\beta}^{TSLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y = \dots$

$$\begin{aligned}\hat{\beta}^{TSLS} &= (\hat{X}'\hat{X})^{-1}\hat{X}'y = \\ &= ((Z\hat{\Pi})'(Z\hat{\Pi}))^{-1}(Z\hat{\Pi})'y = \\ &= ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'y\end{aligned}$$

Двухшаговый МНК

Свойства оценок

Еще раз формула:

$$\hat{\beta}^{TSLS} = ((Z(Z'Z)^{-1}Z'X)'Z(Z'Z)^{-1}Z'X)^{-1}(Z(Z'Z)^{-1}Z'X)'y$$

Состоятельность:

$$\hat{\beta}^{TSLS} \xrightarrow[n \rightarrow \infty]{p} \beta$$

Асимптотическая нормальность:

$$\sqrt{n}(\hat{\beta}^{TSLS} - \beta) \xrightarrow[n \rightarrow \infty]{d} N(\mathbf{0}_{k+1}, \mathbf{V}_{\hat{\beta}^{TSLS}})$$

Ковариационная матрица ошибок:

$$\mathbf{V}_{\hat{\beta}^{TSLS}} = (\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}'_{XZ})^{-1}(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{V}_{zu}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}'_{XZ})(\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}'_{XZ})^{-1}$$