

ЭНДОГЕННОСТЬ И МЕТОД IV. ОЦЕНКА ОТДАЧИ ОТ ОБРАЗОВАНИЯ

Преподаватель: Лукьянова Анна Львовна

К.э.н., доцент Департамента прикладной экономики НИУ ВШЭ,
с.н.с. Центра трудовых исследований НИУ ВШЭ

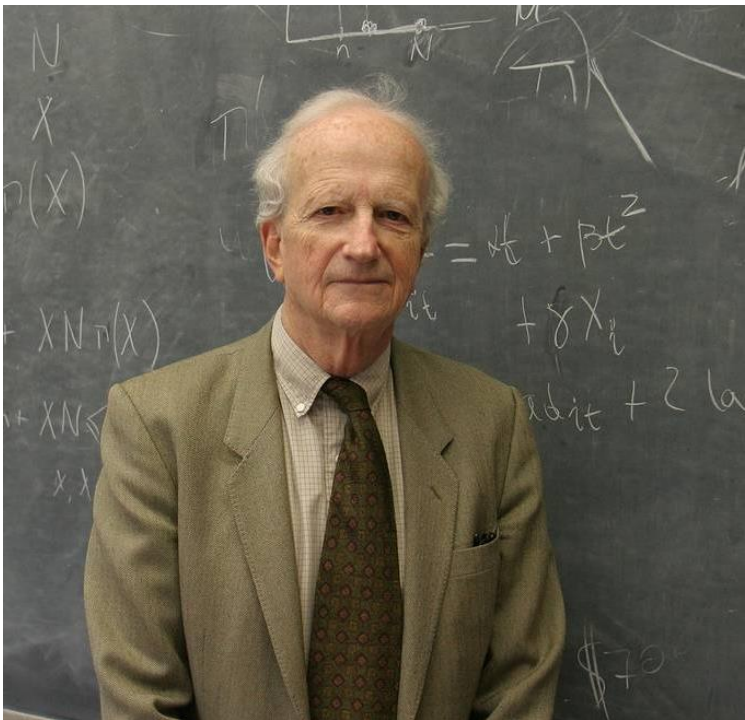
E-mail:

alukyanova@hse.ru

ПЛАН ЛЕКЦИИ

- ❖ Теория человеческого капитала
- ❖ Уравнение Минцера
- ❖ Корреляция и каузальность?
- ❖ Эндогенность. Проблемы МНК
- ❖ Альтернативные методы оценки отдачи от образования:
 - Прокси-переменные
 - Метод инструментальных переменных
 - «Естественные» эксперименты

1



Гэри Беккер (1930-2014)

Лауреат Нобелевской премии 1992 г.

1962 г. – статья «Инвестиции в человеческий капитал: теоретический анализ»,

1964 г. – книга «Человеческий капитал: теоретический и эмпирический анализ».

ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА



Теодор Шульц (1902-1998)

Лауреат Нобелевской премии 1979 г.

Schultz T.W. Investment in Human Beings. — Chicago: University of Chicago Press, 1962.

Schultz T.W. The Economic Value of Education. — New York: Columbia University Press, 1963

ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА

- **Человеческий капитал** – воплощенный в человеке запас способностей, знаний, навыков и мотиваций.
- ЧК накапливается через:
 1. **формальное образование**,
 2. накопление опыта в процессе работы
 3. переподготовку на производстве.
- **Образование** – это не столько потребительское благо, а **инвестиции** (!), которые связаны с **издержками** и приносят дополнительные **доходы** (за счет более высокой производительности труда)
- Издержки: прямые (стоимость обучения, затраты на учебники...), косвенные (утраченные заработки), психологические
- Решение о продолжении обучения принимается на основе сравнения издержек и доходов (с учетом дисконтирования)

ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА: ФОРМАЛЬНАЯ МОДЕЛЬ (1)

Пусть текущая стоимость \$y\$, выплаченных через 1 год равна:

$$PV_1 = \frac{y}{1+r}$$

тогда текущая стоимость \$y\$, выплаченных через 2 года:

$$PV_2 = \frac{y}{(1+r)^2}$$

текущая стоимость \$y\$, выплаченных через T лет:

$$PV_T = \frac{y}{(1+r)^T}$$

где r — ставка дисконтирования

ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА: ФОРМАЛЬНАЯ МОДЕЛЬ (2)

Если человек заканчивает только среднюю школу, то приведенная стоимость его доходов равна:

$$PV_{SCH} = w_{SCH} + \frac{w_{SCH}}{(1+r)} + \frac{w_{SCH}}{(1+r)^2} + \dots + \frac{w_{SCH}}{(1+r)^{46}}$$

Если человек заканчивает университет (4 года), то приведенная стоимость его доходов равна:

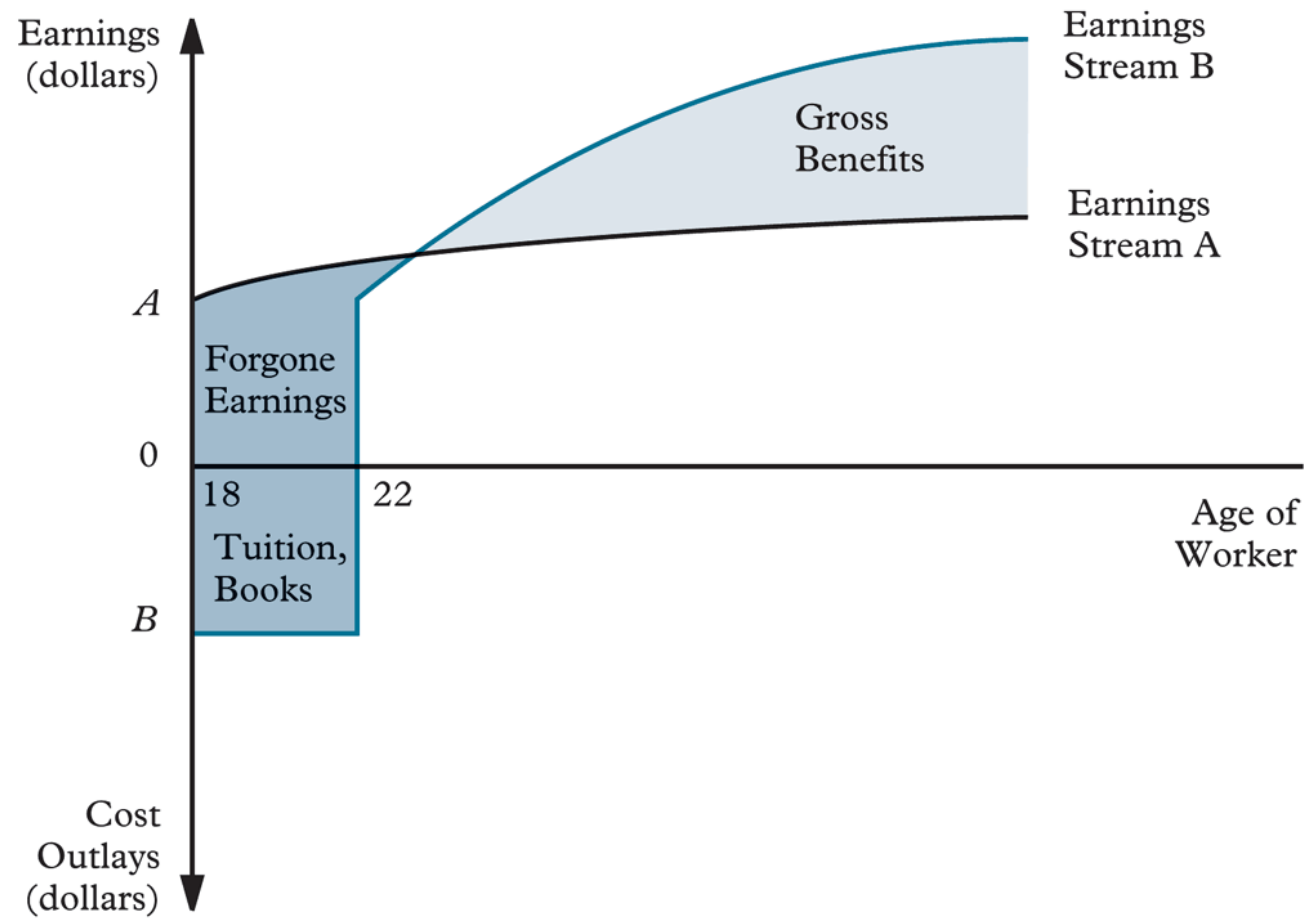
$$PV_{UNI} = -C_{UNI} - \frac{C_{UNI}}{(1+r)} - \frac{C_{UNI}}{(1+r)^2} - \frac{C_{UNI}}{(1+r)^3} + \frac{w_{UNI}}{(1+r)^4} + \frac{w_{UNI}}{(1+r)^5} + \dots + \frac{w_{UNI}}{(1+r)^{46}}$$

Решение о поступлении в университет (инвестиции в высшее образование) оправданно, если:

$$PV_{UNI} > PV_{SCH}$$

При условии, что человек не совмещает работу и учебу.

ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА: ФОРМАЛЬНАЯ МОДЕЛЬ (3)



ТЕОРИЯ ЧЕЛОВЕЧЕСКОГО КАПИТАЛА



Джейкоб Минцер (1922-2006)

Первый предложил термин человеческий капитал в своей статье «Инвестиции в человеческий капитал и персональное распределение дохода» (1958 г.)

Его именем названо уравнение Минцера, представляющие оценки отдачи от инвестиций в человеческий капитал.

УРАВНЕНИЕ МИНЦЕРА: ТЕОРИЯ

Для больших T приведенная стоимость заработков за всю жизнь для индивидов, учившихся s лет, равна $V(s)$ при нулевых прямых издержках:

$$V(s) = Y(s) \int_s^T e^{-rt} dt = \frac{Y(s)}{r} (e^{-rs} - e^{-rT})$$

Где $Y(s)$ – годовой заработок.

После взятия логарифмов и приведения:

$$\ln Y(s) = \ln Y(0) + rs + \ln((1 - e^{-rT}) / (1 - e^{-r(T-s)}))$$

$\ln((1 - e^{-rT}) / (1 - e^{-r(T-s)})) = 0$ при больших T .

Оптимальный уровень инвестиций – обучение в течение еще одного года имеет смысл до тех пор пока отдача от образования не будет равна ставке дисконтирования r .

УРАВНЕНИЕ МИНЦЕРА: ФОРМУЛА

$$\ln W(S, x) = \alpha + \rho S + \beta_0 x + \beta_1 x^2 + \varepsilon$$

Где S – продолжительность обучения (в годах),

x – продолжительность общего трудового стажа (в годах).

ρ – средняя отдача от образования, показывает на сколько % в среднем возрастает заработная плата с каждым дополнительным годом обучения.

Предложенный метод оценивания – МНК.

УРАВНЕНИЕ МИНЦЕРА: КОРРЕЛЯЦИЯ ИЛИ КАУЗАЛЬНОСТЬ?

Отдача от образования: альтернативные объяснения

■ Теория человеческого капитала

Образование дает знания и навыки, которые реально повышают производительность труда.

➔ Между образованием и заработной платой **ЕСТЬ** причинно-следственная связь

■ Теория образовательных сигналов (signaling theory)

Для работодателя важны способности работника, от которых в решающей степени зависит его производительность. Образование не повышает производительность работников, а лишь «сигнализирует» о способностях, поскольку издержки на образование работника тем ниже, чем он способнее (а значит, и производительнее).

➔ Между образованием и заработной платой **НЕТ** причинно-следственной связи, только корреляция; на самом деле обе переменные зависят от третьей – (ненаблюдаемых) способностей.

УРАВНЕНИЕ МИНЦЕРА: ПРОБЛЕМЫ МНК

Возможно смещение в коэффициенте при переменной образования:

- **Различия в уровне способностей** (ability bias) – частный случай смещений, связанных с проблемой невключения в уравнение значимых переменных (omitted variable bias) – **отдача от образования ЗАВЫШЕНА**
- **Ошибки измерения:** если уровень образования измерен с ошибками \Rightarrow смещение коэффициентов в сторону нуля (attenuation bias) – **отдача от образования ЗАНИЖЕНА**

В обоих случаях возникает проблема ЭНДОГЕННОСТИ

ПРИМЕР: ABILITY BIAS В УРАВНЕНИИ МИНЦЕРА

Истинная модель:

$$(1) \quad Y_i = \alpha + \rho S_i + \gamma A_i + v_i$$

Y_i = логарифм заработной платы, S_i = продолжительность обучения в годах, A_i = индивидуальные способности.

Как правило, информация о способностях отсутствует. Вместо истинной модели оценивается сокращенная регрессия, в которой способности входят в состав ошибки :

$$(2) \quad Y_i = \alpha + r S_i + \eta_i$$

где $\eta_i = \gamma A_i + v$ и ковариация между образованием и истинной ошибкой $\text{Cov}(S, v)=0$, т.е. способности – это единственный источник корреляции между S и ошибкой в истинной модели

ПРИМЕР: ABILITY BIAS В УРАВНЕНИИ МИНЦЕРА

Оценка МНК для коэффициента ρ :

$$\hat{r}_{OLS} = \frac{Cov(Y,S)}{Var(S)}$$

Подставляя истинную модель получаем:

$$\hat{r}_{OLS} = \frac{Cov([\alpha + \rho S + \gamma A + v], S)}{Var(S)}$$

Математическое ожидание равно:

$$E[\hat{r}_{OLS}] = \rho + \gamma \frac{Cov(A,S)}{Var(S)}$$

где $\frac{Cov(A,S)}{Var(S)}$ - коэффициент регрессии из уравнения зависимости A от S

Направление смещения положительное (коэффициент завышен):

- $\gamma > 0$, способности имеют самостоятельный (+) эффект на ЗП
- $\frac{Cov(A,S)}{Var(S)} > 0$, более способные индивиды имеют более высокий уровень образования, поскольку им легче учиться

ABILITY BIAS: ВОЗМОЖНЫЕ РЕШЕНИЯ

- Рандомизированные (случайные) эксперименты
- Прокси-переменные для ненаблюдаемых способностей (н-р, IQ)
- Исследования близнецов
- Инструментальные переменные. Примеры:
 - Образование родителей
 - Естественные эксперименты

ABILITY BIAS: РАНДОМИЗИРОВАННЫЕ ЭКСПЕРИМЕНТЫ

Примеры:

Banerjee, A., Cole, S., Duflo, E., Linden, L. (2007) "Remedying Education: Evidence from Two Randomized Experiments in India", *Quarterly Journal of Economics*, 122(3): 1235-1264.

Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M. (2002) "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment", *American Economic Review*, 92(5): 1535–1558.

Heckman, J., Moon, S., Pinto, R., Savelyev, P., Yavitz, A. (2009) "The Rate of Return to the High/Scope Perry Preschool Program", NBER Working Paper No. 15471.

ABILITY BIAS: КОГНИТИВНЫЕ ТЕСТЫ (IQ)

Идея – использовать IQ в качестве прокси-переменной для способностей

$$(1) \quad \ln(Wage) = \beta_0 + \beta_1 S + \beta_2 Abil + \varepsilon$$

$$(2) \quad IQ = \delta_0 + \delta_1 Abil + u$$

IQ может использоваться в качестве прокси *iff* IQ зависит только от способностей, но не зависит от образования (т.е. действительно измеряет врожденные способности).

$$E(IQ | S, Abil) = E(IQ | Abil) = \delta_0 + \delta_1 IQ$$

Тогда, подставляя (2) в (1), имеем:

$$\ln(Wage) = \left(\beta_0 - \frac{\beta_2 \delta_0}{\delta_1}\right) + \boxed{\beta_1} S + \frac{\beta_2}{\delta_1} IQ + \left(\varepsilon - \frac{\beta_2}{\delta_1} u\right) = \alpha_0 + \beta_1 S + \alpha_2 IQ + e$$

В результате, получаем состоятельную оценку для α_0 , β_1 и α_2 .

Если IQ зависит и от способностей и образования, то имеем: $IQ = \delta_0 + \delta_1 Abil + \delta_2 S + u$ (2')

$$\ln(Wage) = \left(\beta_0 - \frac{\beta_2 \delta_0}{\delta_1}\right) + \boxed{\left(\beta_1 - \frac{\beta_2 \delta_2}{\delta_1}\right)} S + \frac{\beta_2}{\delta_1} IQ + \left(\varepsilon - \frac{\beta_2}{\delta_1} u\right)$$

ABILITY BIAS: ИССЛЕДОВАНИЯ БЛИЗНЕЦОВ

Идея – способности определяются генетически \Rightarrow у монозиготных близнецов они должны быть одинаковы + одинаковое влияние семейного окружения.

Данные - выборка монозиготных близнецов

Оценивается уравнения вида: $Ln(Wage_{1i}) = \alpha X_i + \beta Z_{1i} + \mu_i + \varepsilon_{1i}$

$$Ln(Wage_{2i}) = \alpha X_i + \beta Z_{2i} + \mu_i + \varepsilon_{2i}$$

где $Ln(Wage_{1i})$ и $Ln(Wage_{2i})$ - лог-ЗП 1-го и 2-го близнецов,

X – переменные, одинаковые для обоих близнецов - возраст, раса,

Z – переменные, которые могут отличаться у близнецов, включая образование

μ – ненаблюдаемые семейные характеристики,

ε – ненаблюдаемые индивидуальные характеристики.

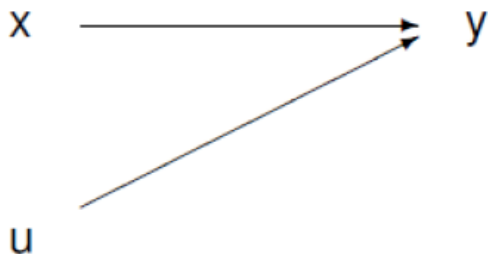
Разность этих двух уравнений:

$$\Delta \ln(Wage) = \beta (Z_{1i} - Z_{2i}) + \varepsilon_{1i} - \varepsilon_{2i}$$

КОГДА И ЗАЧЕМ ИСПОЛЬЗУЕТСЯ IV?

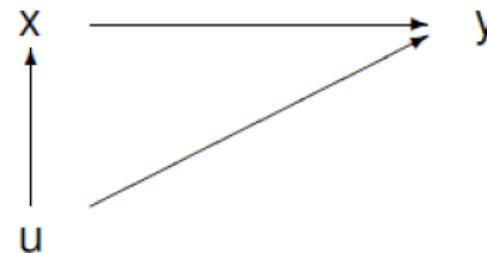
IV используется для преодоления проблемы **эндогенности**, т.е. для того, чтобы получить состоятельные оценки коэффициентов в тех случаях, когда регрессор X коррелирует со случайной ошибкой.

Нормальная ситуация: $y=bX+u$
 X и u не коррелированы. МНК дает
несмещенные и состоятельные оценки



Можно использовать МНК

Эндогенность: $y=bX+u$
 X и u коррелированы. МНК дает
смещенные и несостоятельные оценки



Нельзя использовать МНК:
увеличение числа наблюдений и
добавление регрессоров не спасет МНК

КОГДА И ЗАЧЕМ ИСПОЛЬЗУЕТСЯ IV?

Типичные ситуации, когда возникает эндогенность:

1. Самоотбор
2. Пропущенные переменные, коррелированные с используемыми регрессорами (например, способности в уравнении заработной платы)
3. Ошибки измерения в регрессорах
4. Обратная причинность и одновременность – наличие одновременной зависимость между Y и X и между X и Y (например, цена и количество в модели спроса и предложения; число полицейских и уровень преступности; заболеваемость и число врачей)

СУТЬ МЕТОДА IV

Представим, что вариация эндогенной переменной X состоит из двух частей:

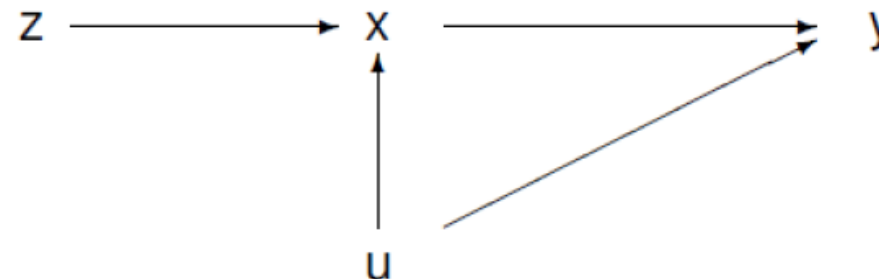
1. Одна часть по каким-то причинам коррелирует с ошибкой
2. Другая часть не коррелирует с ошибкой

Задача – **изолировать ту часть вариации X , которая не коррелирует с ошибкой и использовать только ее при оценке коэффициента при X . Устранив из уравнения ту часть вариации X , которая коррелирует с ошибкой, мы убираем смещение в МНК**

Переменная Z (инструмент) несет информацию о том, какая часть X не коррелирована с ошибкой. Для этого Z должна быть:

1. (*instrument relevance*) коррелирована с X : $\text{Corr}(Z, X) \neq 0$
2. (*instrument exogeneity, exclusion restriction*) не коррелирована с ошибкой u : $\text{Corr}(Z, u) = 0$, т.е. не имеет самостоятельного влияния на y кроме как через X

Оба условия чрезвычайно важны, чтобы Z была хорошим инструментом!!!



ДВУХШАГОВЫЙ МНК, 2-МНК (TWO-STAGE LEAST SQUARED, 2SLS)

1. Построить дополнительную регрессию (first stage equation):

$$S_i = X_i' \pi_{10} + \pi_{11} Z_i + \xi_{1i}$$

Оценить ее, используя МНК. Если в уравнении для Y были другие экзогенные переменные, то их также нужно включить в уравнение для S .

2. Рассчитать из оцененного уравнения, прогнозные значения для \hat{S} .

$$\hat{S}_i = X_i' \hat{\pi}_{10} + \hat{\pi}_{11} \tilde{Z}_i$$

3. Подставить \hat{S} вместо S в исходное уравнение и оценить его, используя МНК (second stage):

$$Y_i = X_i' \alpha + \rho \hat{S}_i + error$$

Не стоит делать эту процедуру вручную. Чтобы получить правильные стандартные ошибки используйте специальные команды в Стате: `ivregress`

Кроме 2МНК можно использовать обобщенный метод моментов (GMM), метод максимального правдоподобия с ограниченной информацией (LIML).

IV В STATA

ivregress **estimator** depvar [varlist1] (varlist2 = varlist_iv) [if...] [, options]

Estimator (метод оценивания): 2sls, liml, gmm

varlist1 – список экзогенных регрессоров

varlist2 – список эндогенных регрессоров

varlist_iv – список инструментов

Пример: ivregress 2sls y x1 x2 (y2 y3 = z3 z4 z5), small

Заметьте, что инструментами как для y2, так и для y3 являются все инструменты z3 z4 z5 + экзогенные переменные x1 x2

Более продвинутая команда для IV (ado-file): ivreg2

ПРОВЕРКА ВАЛИДНОСТИ ИНСТРУМЕНТОВ

Хороший инструмент должен удовлетворять 2 условиям:

1. (*instrument relevance*) Z коррелирован с X: $\text{Corr}(Z, X) \neq 0$

2. (*instrument exogeneity, exclusion restriction*) Z не коррелирован с ошибкой u : $\text{Corr}(Z, u) = 0$

Условие 1 (релевантность):

Чем больше вариации в X объясняют инструменты, тем больше информации используется в IV регрессии → тем точнее оценки.

Инструменты, не удовлетворяющие условию 1, называются *слабыми*.

- **Полезное правило №1:** F-stat по инструментам (`estat firststage`) в уравнении на 1-м шаге д.б. > 10
- **Полезное правило №2:** При сомнениях в качестве инструментов для оценивания рекомендуют использовать метод (limited information ML) LIML, который на малых выборках дает меньшее смещение, чем 2SLS и GMM. Но тоже с осторожностью

Тесты после `ivregress` в STATA:

`estat endogenous; estat firststage`

ПРОВЕРКА ВАЛИДНОСТИ ИНСТРУМЕНТОВ

Условие 2. Экзогенность инструментов – $\text{Corr}(Z, u=0)$:

1. Если N инструментов равно N эндогенных переменных (модель точно идентифицирована) – формальный тест невозможен. Но возможен тест Хаусмана на сравнение МНК и IV.
2. Если N инструментов больше N эндогенных переменных (модель сверхидентифицирована) – возможен тест на сверхидентифицирующие ограничения (*overidentifying restrictions*) – тест Саргана после 2МНК, Hansen's J – тест после GMM.

Тест Саргана, J-test (`estat overid`) – строится регрессия остатков IV регрессии на все использованные инструменты.

Нулевая гипотеза – остатки не коррелированы с инструментами. Если H_0 отклоняется, то один или более инструментов коррелированы остатками. К сожалению, невозможно определить какой из инструментов

Тесты после `ivregress` в STATA:

`estat overid`

ЭНДОГЕННОСТЬ И ИНСТИТУТЫ

Естественные эксперименты (“Natural experiments”) – ситуации, когда силы природы, шоковые воздействия или меры государственной политики создают условия, близкие к условиям рандомизированного эксперимента.

В соответствии с теорией ЧК, люди принимают образовательные решения на основе сравнения выгод и издержек дополнительного образования.

→ В случае отдачи от образования инструменты могут быть 2 видов:

1. Различия в стоимости обучения, например из-за различий в условиях предоставления образовательных кредитов и субсидий (не зависящие от уровня способностей), из-за географических и т.п. объективных различий
2. Различия в институциональных характеристиках, определяющих продолжительность обучения.

ЭНДОГЕННОСТЬ И ИНСТИТУТЫ: ВРЕМЯ РОЖДЕНИЯ (1)

Angrist & Krueger (1991)

Во многих штатах закон требует, чтобы дети поступали в школу в год, когда им исполняется 6 лет. Тогда, родившиеся в 4 кв. поступают в школу в возрасте $5\frac{3}{4}$, а родившиеся в 1 кв. в возрасте $6\frac{3}{4}$. В то же время, закон об обязательном образовании требует, чтобы дети продолжали обучение в школе до достижения 16 лет. В результате, на момент достижения 16 лет, ученики, родившиеся в 1 и 4 кв., будут учиться в разных классах. Эти два законодательных требования создают условия для появления «естественного эксперимента»

Figure 1

Mean Years of Completed Education, by Quarter of Birth

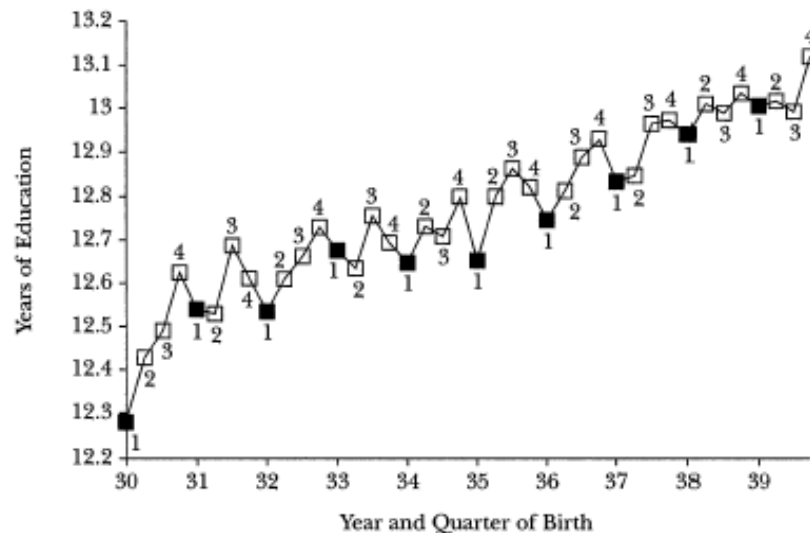
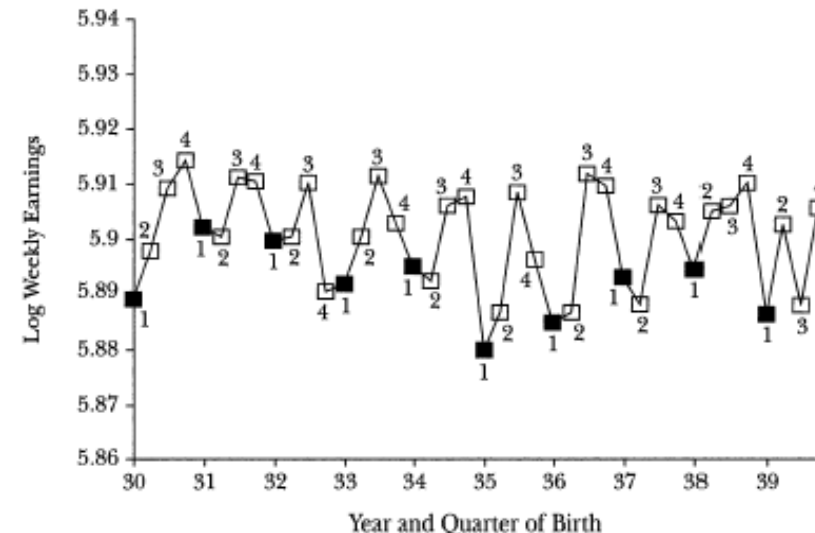


Figure 2

Mean Log Weekly Earnings, by Quarter of Birth



ЭНДОГЕННОСТЬ И ИНСТИТУТЫ: КВАРТАЛ РОЖДЕНИЯ (2)

Angrist & Krueger (1991)

Квартал рождения коррелирует с продолжительностью обучения, но не зависит от природных способностей, мотивации к труду, семейных характеристик → время рождения м.б. хорошим инструментом для переменной числа лет обучения.

Т.о. в данном случае для идентификации отдачи на образование используется лишь небольшая часть вариации в продолжительности обучения – та часть, которая определяется временем рождения: в среднем люди родившиеся в 1 кв учатся на 0,1 года меньше и зарабатывают на 0,1% меньше, чем те, кто родился в другом квартале

Результаты: оценки выше, чем полученные по МНК примерно на 30%!

МНК И IV С ИНСТИТУЦИОНАЛЬНЫМИ ИНСТРУМЕНТАМИ: СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Author	Sample and Instrument		Schooling Coefficients	
			OLS	IV
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and indicators for race, marital status, urban residence.	1920–29 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930–39 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940–49 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth. LIML estimates.	1930–39 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940–49 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience. Note: Schooling measured in units of college credit equivalents.	Models without test score or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) Instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with parental education. Controls include race, experience (treated as endogenous), region, and parental education	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)
		Models that use college proximity × family background as instrument	—	0.097 (0.048)

ПОЧЕМУ IV-ОЦЕНКИ ОТДАЧИ НА ОБРАЗОВАНИЕ ВЫШЕ, ЧЕМ МНК-ОЦЕНКИ?

1. Griliches (1977): IV-оценки нивелируют (негативное) смещение, связанное с ошибками измерения, в то время как (положительное) смещение, вызванное способностями, достаточно мало.
2. Bound & Jaeger (1996): Между контрольной группой и группой, подвергшейся воздействию, имеются ненаблюдаемые различия, которые смещают IV-оценки вверх
3. Ashenfelter & Harmon (1998): “publication bias” – отбирая спецификации для публикации, исследователи отдают предпочтение статистически значимым IV-оценкам
4. Card (1999): IV-оценки относятся к группам с более низким уровнем образования и более высокой предельной нормой отдачи на образование

ПОЧЕМУ МЫ РЕДКО ИСПОЛЬЗУЕМ IV?

Cameron & Trivedi (2005) Microeconometrics

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

Проблема находений хороших инструментов: многие переменные влияют не только на эндогенную переменную, но и имеют прямое влияние на зависимую переменную.

Оценки IV всегда являются смещенными (хотя и состоятельными). Все обоснование базируется на асимптотике, свойства оценок для конечных выборок не столь хороши. Поэтому их использование для малых выборок крайне проблематично.

Эффективность IV-оценок ниже чем в МНК. При наличии слабых инструментов (имеющих низкую корреляцию с эндогенными переменными) потери в точности, могут быть столь значительны, что IV-оценки не дают преимуществ перед МНК

Зачастую оцененный эффект может интерпретироваться только как LATE

СЛАБЫЕ ИНСТРУМЕНТЫ (WEAK INSTRUMENTS)

- Длительное время исследователей волновал вопрос о поведении оценок на малых выборках
- В начале 1990-х годов появилось несколько статей, в которых указывалось, что основным источником смещения являются слабые инструменты – те инструменты, которые слабо коррелированы с эндогенной переменной + ситуации, когда используется много инструментов для одной эндогенной переменной
- Если корреляция слабая, то оценки IV смещаются в сторону МНК-оценок, но с большими стандартными ошибками.

СЛАБЫЕ ИНСТРУМЕНТЫ

Если истинная модель имеет вид:

$$y = \beta x + \eta$$

Уравнение первого шага:

$$x = \mathbf{Z}'\pi + \xi$$

где \mathbf{Z} – вектор инструментов.

Смещение, связанное с IV, равно:

$$E[\hat{\beta}_{SLS} - \beta] \approx \frac{\sigma_{\eta\xi}}{\sigma_{\xi}^2} \frac{1}{F+1}$$

Смещение в МНК

где F – F -статистика по инструментам в уравнении первого шага.

Если $F \rightarrow 0$, то смещение стремится к смещению в МНК-регрессии

Если $F \rightarrow \infty$, то смещение стремится к нулю

ВЫБОР ИНСТРУМЕНТОВ

1. Инструментами могут быть как непрерывные, так и бинарные переменные.
2. Число инструментов должно быть не меньше, чем число эндогенных переменных. Но не всегда чем больше инструментов, тем лучше (см. пункт 5)
3. Помнить, что инструменты не должны иметь самостоятельного влияния на Y кроме как через эндогенную переменную.
4. Осторожно применять IV на малых выборках (оценки IV остаются смещенными на конечных выборках, хотя и состоятельными, и при малых N смещение может быть значительным).
5. Остерегайтесь слабых (weak) инструментов, т.е. тех, которые слабо коррелированы с X . Смещения при использовании слабых инструментов могут быть больше, чем при МНК.