

2013年8月17日

JAGES研究会

操作変数法

ー基本的アイデアと適用方法

日本福祉大学経済学部

遠藤 秀紀

はじめに

- 【パート1】基本的な考え方
 - 【パート2】適用の流れ
 - 【パート3】データを用いた実習
 - An Introduction to Modern Econometrics Using Stata
 - Microeconometrics Using Stata
-

【パート1】基本的な考え方

1. 操作変数法(IV)は、どんな状況で使うのか？
2. IV(確認)
 - 2-1. 推定の考え方
 - 2-2. 推定後のチェックポイント
(操作変数に求められる条件)

1. IVを使う状況

1. 操作変数法(IV)は、どんな状況で使うのか？

注目する独立変数 X が

観測できない情報と相関する場合

└──────────→ 誤差項 e

(a) X に測定誤差がある場合

(b) 重要なコントロール変数が得られない場合
(除外変数の問題)

2-1. 通常の回帰イメージ

(例) 就労(L)は健康(H)によいのか？

$$H = X\alpha + \beta L$$

$$\beta > 0$$



観測できない情報との相関

(例) 就労(L)は健康(H)によいのか？

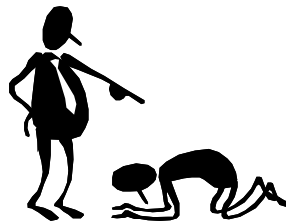
$$H = X\alpha + \beta L$$

$\beta > 0?$



人間関係

厄介な交渉



観測できない情報の影響

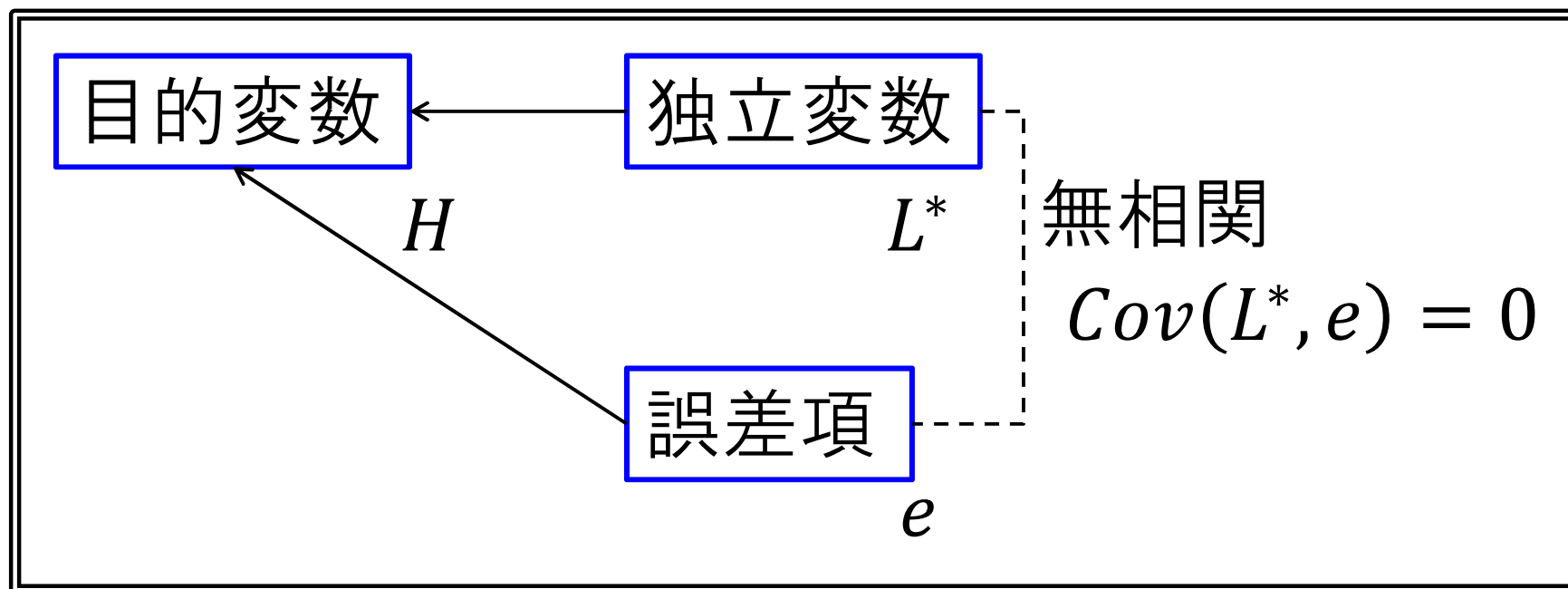
通常の回帰式(最小二乗法など)

目的変数 $\longrightarrow$ $H = X\alpha + \beta L^* + e$

独立変数 L^* (外生変数)

誤差項 e (誤差項と相関しない)

偏りのない推定値: $\hat{\beta} = \beta$

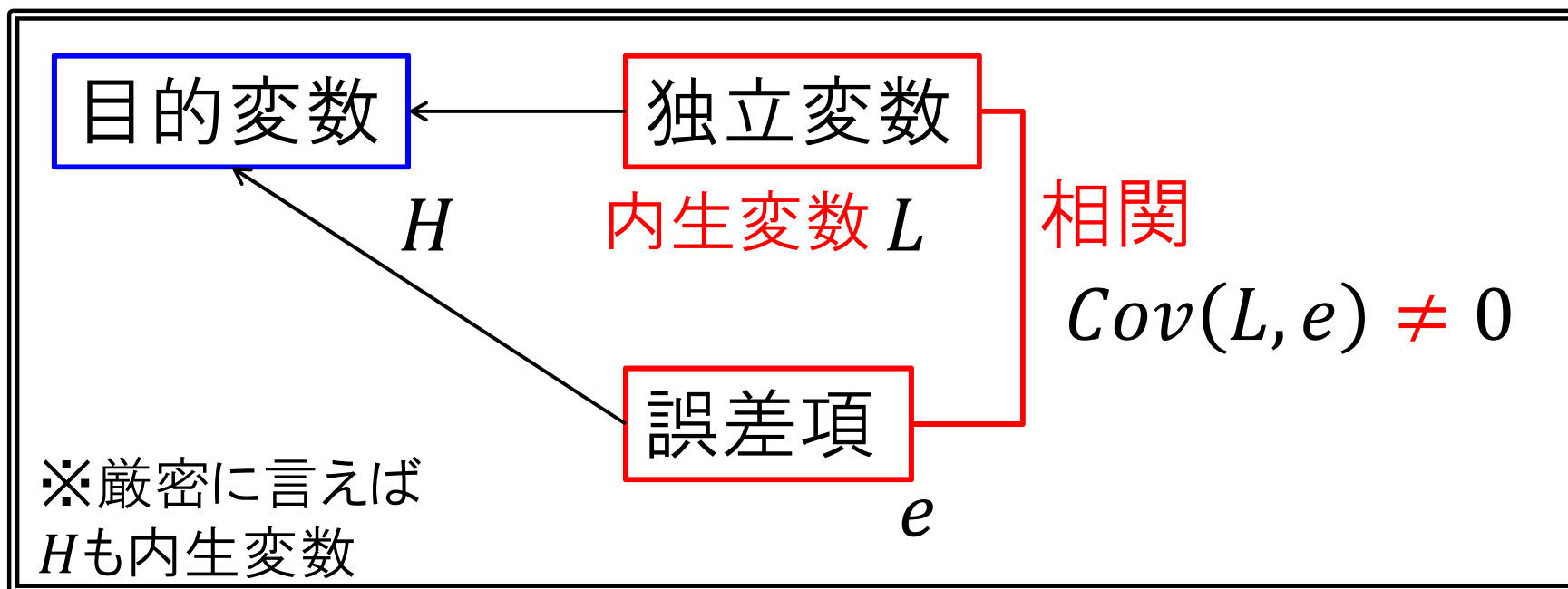


独立変数と誤差項が相関する場合

独立変数 内生変数

誤差項 e と相関する
 $e \rightarrow e(L)$

$$H = X\alpha + \beta \boxed{L} + \boxed{e}$$

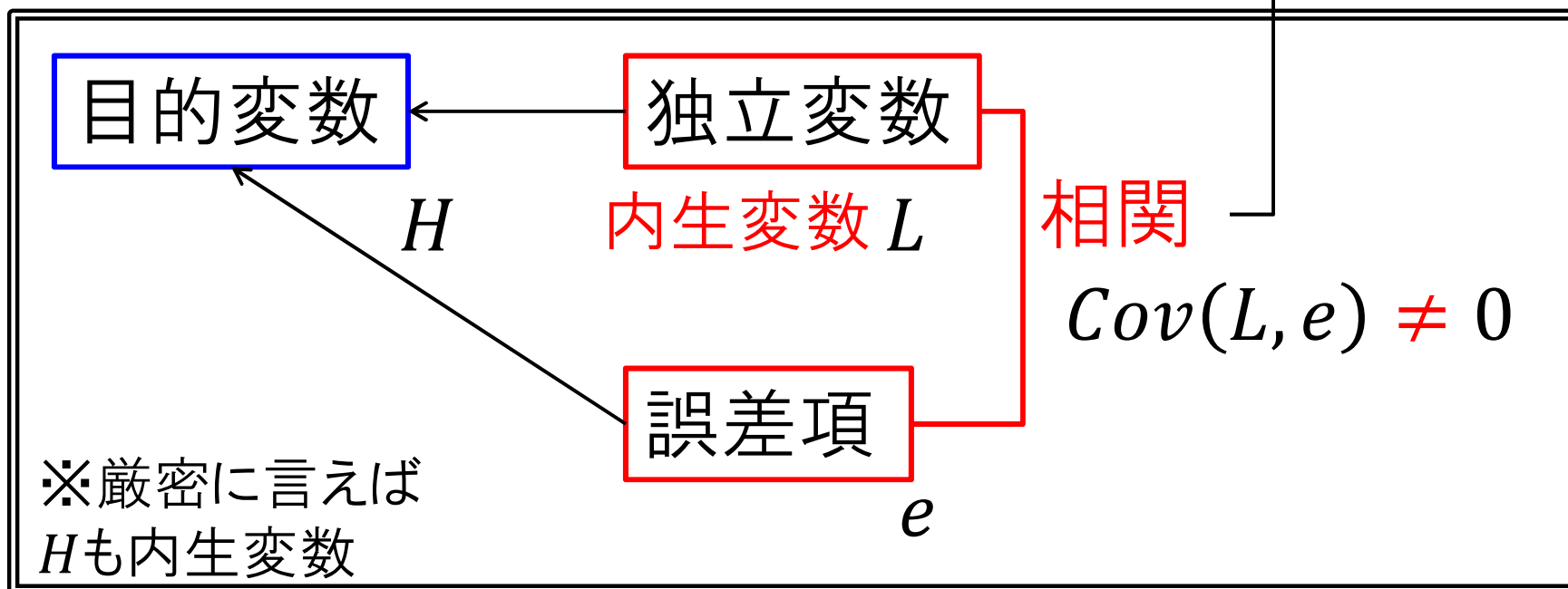


独立変数と誤差項が相関する場合

$$H = X\alpha + \beta L + e$$

偏りのある推定値: $\hat{\beta} = \beta + \frac{de(L)}{dL}$ ← ゼロにならない = 偏り

(参考) L が二値変数のとき: $\hat{\beta} = \beta + (e_{L=1} - e_{L=0})$



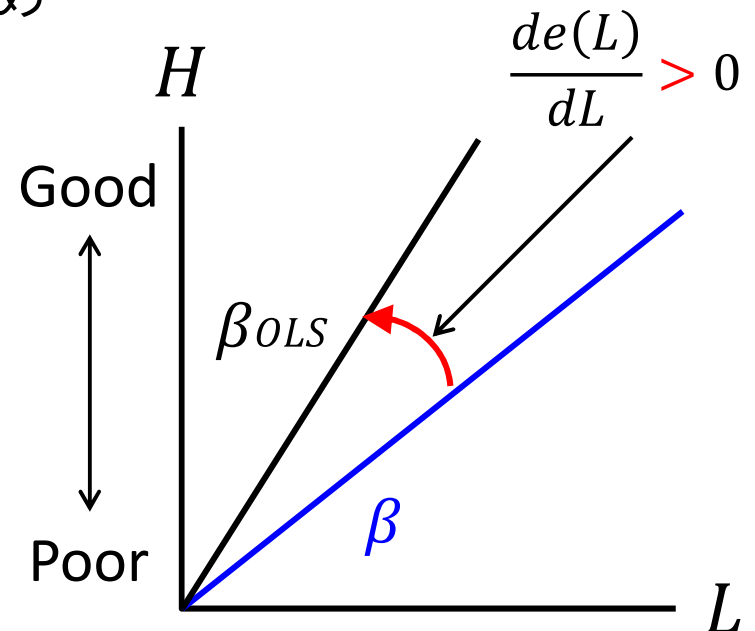
推定値の偏り(就労と健康の例)

$$\frac{de(L)}{dL} > 0 \rightarrow \beta_{OLS} > \beta$$

就労を通じて H を増加 β_{OLS} は上方に偏る

○過去、健康だった人が働いているため

- Snyder and Evans (2006)
- Mandal and Roe (2008)



推定値の偏り(就労と健康の例)

$$\frac{de(L)}{dL} < 0 \rightarrow \beta_{OLS} < \beta$$

就労を通じて H を減少 β_{OLS} は下方に偏る

(a)職場環境への適応

Marmot et al. (1999)

Karasek (1979)「仕事要求度－裁量度」モデル

OECD (2005) 相対的な職業訓練投資額の減少(アメリカ)

厚生労働省(2002)

(b)希望の雇用形態・就業時間と実態のズレ

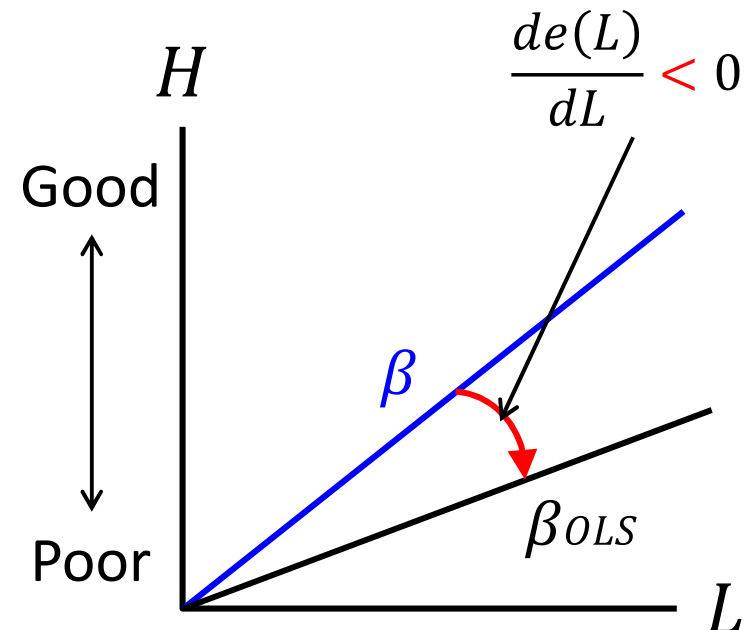
高木(2008), OECD (2006)

(c)生産性の低下

Hellerstein and Neumark (2007), OECD (2006)

(d)その他

起業した事業の不振、後継者不足による就業期間延長

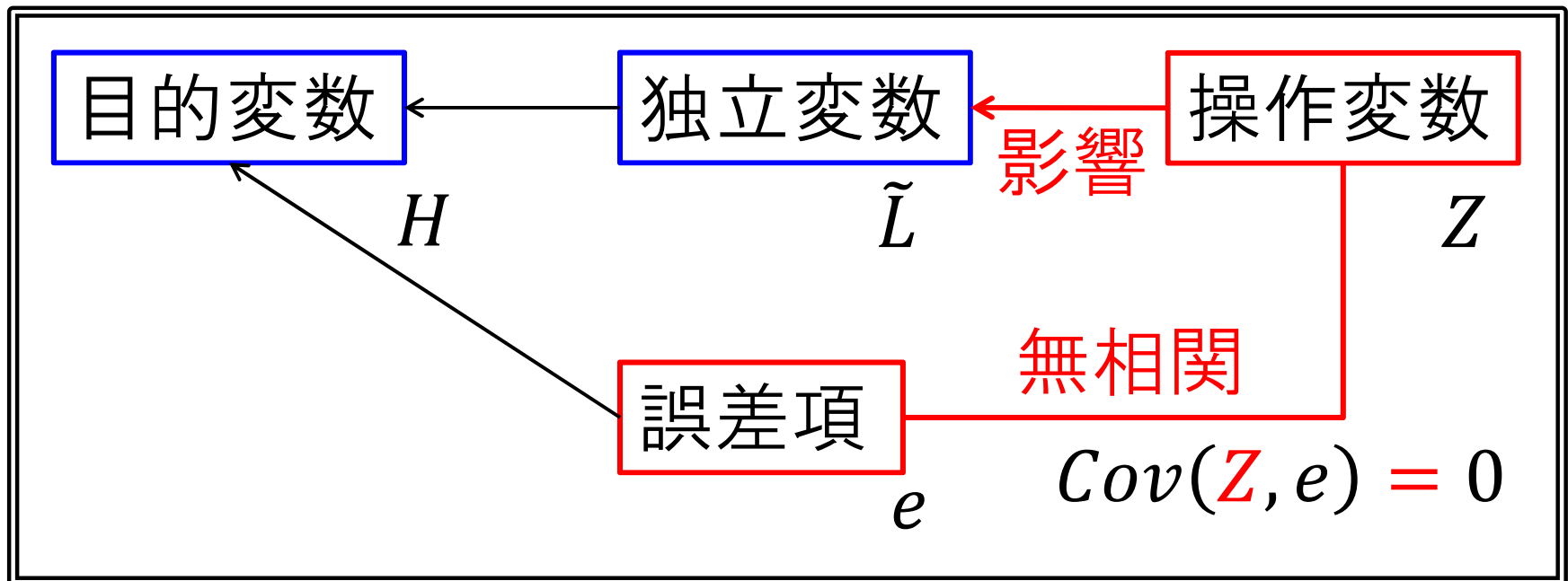


操作変数法(IV)とは

誤差項と相関する独立変数(内生変数) L から
「誤差項 e と相関しない部分」を抽出して、推定値の偏りを除く方法



操作変数(内生変数 L のみに影響し、
誤差項 e と相関しない変数) Z を用いて抽出



(直観的な)IVの手順

$$\begin{cases} H = X\alpha + \beta L + e \\ L = X\gamma + \theta Z + u \end{cases}$$

(第1段階) 操作変数 Z に内生変数 L を回帰

$$L = X\gamma + \theta Z + u$$

(第2段階) L の予測値 $\hat{L}$ に H を回帰

$$H = X\alpha + \beta \hat{L} + e \rightarrow \text{一致推定量}\tilde{\beta}\text{を得る}$$

(注1) 標本数が多くないと、推定量が漸近正規性を満たさない

(注2) Z が妥当なとき、次の関係が確認できる $Cov(Z, e) = 0$

$$Cov(\tilde{L}, e) = 0$$

$$Cov(e, u) \neq 0$$

2-2. 操作変数に求められる条件

(条件1: 識別性)

操作変数(群) Z は、**内生変数 L にのみ影響し**、
もとの推定式 $H = X\alpha + \beta L + e$ の**誤差項 e と無相関**であること

識別性のチェック

(条件2: 強度)

内生変数 L の変動に対する**説明力が弱くないこと**

弱い操作変数のチェック

(付帯状況: 誤差項)

妥当な操作変数が選ばれるとき、
2本の推定式の誤差項 e, u
(もしくは e と H)は無相関でない

内生性のチェック
(内生変数は本当に「内生」か)

識別性のチェック(過剰識別制約の検定)

操作変数 Z と残差 $\hat{e}$ が無相関かどうかを検証

※ Z の数 > 誤差項と相関する独立変数

(過剰識別(Over-identification))

(1) $\hat{e}$ を操作変数法により推定

(2) $\hat{e}$ にすべての外生変数と操作変数を回帰

$$\hat{e} = \mathbf{X}\kappa + \mathbf{Z}\xi + \mu$$

(3) F 値(もしくは R^2)を求める

(4) J 統計量を求める

$$J = \text{標本数} \times F\text{値}$$

(5) J 統計量は、自由度 = 操作変数の数 - 1の χ^2 乗分布に従うことを利用して、検定を行う(H_0 : $\mathbf{Z}$ は $\hat{e}$ と無相関である($\xi = 0$))

(6) 有意水準5%を超える場合、帰無仮説は棄却できない

($\mathbf{Z}$ と $\hat{e}$ の相関は、統計的に確認できない)

弱い操作変数のチェック

Z が内生変数 L の操作変数として機能するか、検証

(イメージ)※内生変数が1つの場合

(1) $L = Z\tau + u$ を計算し、F値を求める

(2) $F > 10$ であれば、弱い操作変数ではない

Staiger and Stock(1997)、Stock, Wright and Yogo (2002)

【パート2】適用の流れ

1. 適用前に考えること

OLSやロジスティック回帰などで済ませられないか
注目する独立変数を「内生」と考えることは重要か
操作変数となりそうなデータを準備できるか

2. 適用を決めてから注意すること

目的変数と内生変数の形状

操作変数の数

操作変数は弱くないか

内生変数と考えた変数は、「内生」と考えてよいか

操作変数となりそうなデータを準備できるか

(ポイント1)

対象(回答者など)の意思決定の範囲外の事象であること

(ポイント2)

内生変数のみに影響し、目的変数 H と

(論理的な)関連が見出し難いこと

※内生変数の発生と同時か、それ以前の事象(が望ましいと思う)

※目的変数(推定式)が異なれば、操作変数も異なる

(ポイント3)

とりあえず、操作変数の候補に目的変数を直接回帰してみる

(候補となりやすい？変数)

自宅の場所、地域ダミー、生年月日(早生まれ)、遺伝子情報、
諸制度の変更、親の属性、Fuzzy Regression Discontinuity Design

時々見られる現象

1. 操作変数が見つからない

操作変数の候補を組み合わせる(交差項など)

t 期前の内生変数を用いる(誤差項の系列相関に要注意)

あきらめる

2. 操作変数の推定値が論理的に矛盾する

他の変数を操作変数の候補として探す

3. 内生性がどうしても確認できない

「この独立変数は内生変数」という見解の間違い

操作変数の選択が良くない

変数の形状と操作変数法

1. 目的変数・内生変数(独立変数)が連続変数

二段階最小二乗法(2SLS) ← 誤差項が*i.i.d.*に従うとき

一般化モーメント法(GMM) ← 誤差項が*i.i.d.*に従わないとき

2. 目的変数が二値・内生変数(独立変数)が連続

Probit model with endogenous covariates

3. 目的変数が打ち切り・内生変数(独立変数)が連続

Tobit model with endogenous covariates

4. 目的変数が連続・内生変数(独立変数)が二値

処置効果モデル(Treatment effect model)

【パート3】データを用いた実習(1)

1. 二段階最小二乗法(2SLS)

「若年層の賃金の決定要因に関する研究」より

(出典) Griliches, Z. (1976) "Wages of Very Young Men", *Journal of Political economy*, 84, S69-S85.

Modified: Baum, C. F. (2006) *An Introduction to Modern Econometrics Using Stata*, Stata Press.

データファイル: griliches.dta

推定式: $\log(wage) = X\alpha + \beta IQscore + e$ (N=758)

目的変数
(賃金(対数))

内生変数(IQテストのスコア)

- ✓ 仕事への適応の柔軟さ
 - ✓ 賃金の上がりやすい業種への就職など、
- 未観測情報との相関が疑われる

【パート3】データを用いた実習(1)

*IQscore*を操作変数に回帰するための式を追加

$$\log(wage) = X\alpha + \beta IQscore + e$$

$$IQscore = X\gamma + Z\theta + u$$

X: 独立変数(外生)

s(教育年数)

expr(卒業後の年数(?))

tenure(在職期間)

rns(南部居住=1)

smsa(標準大都市統計圏=1)

*_lyear_**(調査年(67~73年))

Z: 操作変数(の候補)

med(母親の教育年数)

kww(別の共通試験のスコア)

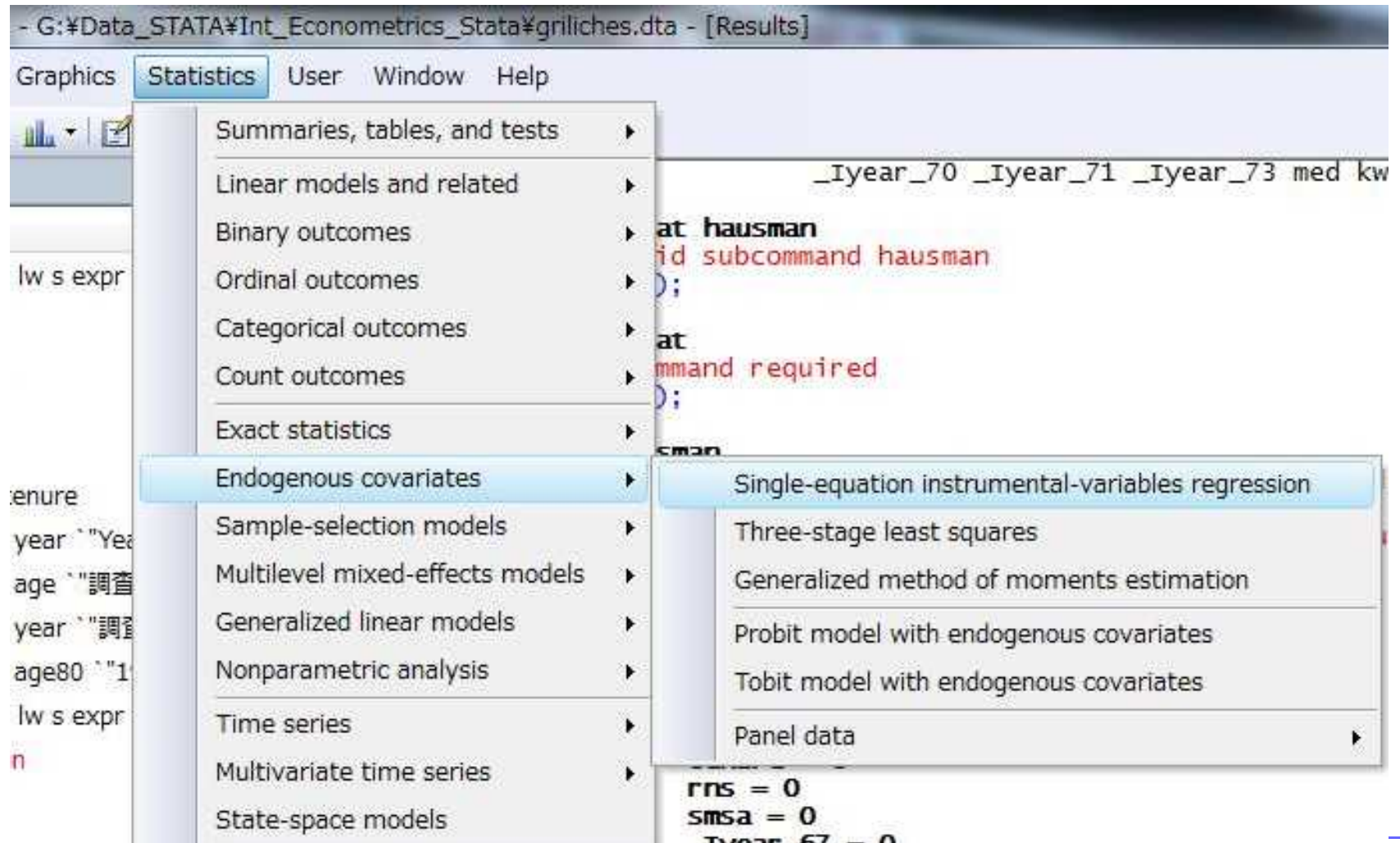
age(年齢)

mrt(結婚している=1)

→ 2SLSをStataで実行しましょう

Stataで2SLSを実行するには

(1) Stataのツールバーから、Statistics → Endogenous covariates
→ “Single-equation instrumental-variables regression”を選択



Stataで2SLSを実行するには

(2)項目を選択or 入力

独立変数(群)

目的変数

内生変数

操作変数(群)

※過剰識別

推定方法

※ここでは2SLS

The screenshot shows the 'ivregress - Single-equation instrumental regression' dialog box. The 'Model' tab is selected. The 'Dependent variable' field contains 'lw'. The 'Independent variables' field contains 's expr tenure rms smsa _lyear_67 _lyear_68 _lyear_69 _lyear_70 _lyear_71'. The 'Endogenous variables' field contains 'iq'. The 'Instrumental variables' field contains 'med kww age mrt'. The 'Estimator' section has three radio buttons: 'Two-stage least squares (2SLS)' (selected), 'Limited-information maximum likelihood (LIML)', and 'Generalized method of moments (GMM)'. The 'Treatment of constant' section has two checkboxes: 'Suppress constant term' and 'Has user-supplied constant'. The 'Reporting' tab is highlighted in the top bar. The bottom of the dialog has 'OK', 'Cancel', and 'Submit' buttons.

※Reportingのタブで

“Report first-stage regression”を選択すると
第1段階・第2段階両方の結果を出力

Stataで2SLSを実行するには

(3)結果がメイン画面に出力(第1段階の推定結果)

First-stage regressions

内生変数を操作変数に
回帰した結果

Number of obs = 758
F(15, 742) = 25.03
Prob > F = 0.0000
R-squared = 0.3360
Adj R-squared = 0.3226
Root MSE = 11.2088

iq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
s	2.497742	.2858159	8.74	0.000	1.936638	3.058846
expr	-.033548	.2534458	-0.13	0.895	-.5311042	.4640082
tenure	.6158215	.2731146	2.25	0.024	.0796522	1.151991
rns	-2.610221	.9499731	-2.75	0.006	-4.475177	-.7452663
smsa	.0260481	.9222585	0.03	0.977	-1.784499	1.836595
_Iyear_67	.9254935	1.655969	0.56	0.576	-2.325449	4.176436
_Iyear_68	.4706951	1.574561	0.30	0.765	-2.620429	3.56182
_Iyear_69	2.164635	1.521387	1.42	0.155	-.8221007	5.15137
_Iyear_70	5.734786	1.696033	3.38	0.001	2.405191	9.064381
_Iyear_71	5.180639	1.562156	3.32	0.001	2.113866	8.247411
_Iyear_73	4.526686	1.48294	3.05	0.002	1.615429	7.437943
med	.2877745	.1622338	1.77	0.077	-.0307176	.6062665
kww	.4581116	.0699323	6.55	0.000	.3208229	.5954003
age	-.8809144	.2232535	-3.95	0.000	-1.319198	-.4426307
mrt	-.584791	.946056	-0.62	0.537	-2.442056	1.272474
_cons	67.20449	4.107281	16.36	0.000	59.14121	75.26776

Stataで2SLSを実行するには

(3)結果がメイン画面に出力(第2段階の推定結果)

Instrumental variables (2SLS) regression

Number of obs = 758
wald chi2(12) = 560.57
Prob > chi2 = 0.0000
R-squared = 0.4255
Root MSE = .32491

内生変数

	lw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	iq	.0001747	.0039035	0.04	0.964	-.007476	.0078253
	s	.0691759	.0129366	5.35	0.000	.0438206	.0945312
	expr	.029866	.0066393	4.50	0.000	.0168533	.0428788
	tenure	.0432738	.0076271	5.67	0.000	.0283249	.0582226
	rns	-.1035897	.029481	-3.51	0.000	-.1613715	-.0458079
	smsa	.1351148	.0266573	5.07	0.000	.0828674	.1873623
	_Iyear_67	-.052598	.0476924	-1.10	0.270	-.1460734	.0408774
	_Iyear_68	.0794686	.0447194	1.78	0.076	-.0081797	.1671169
	_Iyear_69	.2108962	.0439336	4.80	0.000	.1247878	.2970045
	_Iyear_70	.2386338	.0509733	4.68	0.000	.1387281	.3385396
	_Iyear_71	.2284609	.0437436	5.22	0.000	.1427251	.3141967
	_Iyear_73	.3258944	.0407181	8.00	0.000	.2460884	.4057004
	_cons	4.39955	.2685443	16.38	0.000	3.873213	4.925887

Instrumented: iq
Instruments: s expr tenure rns smsa _Iyear_67 _Iyear_68 _Iyear_69
_Iyear_70 _Iyear_71 _Iyear_73 med kww age mrt

ここまでの結果を見ると...

(第1段階) $IQscore = X\gamma + \mathbf{Z}\theta + u$

$\mathbf{Z}$: 操作変数(の候補)

mrt 以外は $P < 0.1$

med の推定値 = 0.288($P = 0.077$)

kww の推定値 = 0.458($P = 0.000$)

age の推定値 = -0.881($P = 0.000$)

mrt の推定値 = -0.585($P = 0.537$)

(第2段階) $\log(wage) = X\alpha + \beta IQscore + e$

$\tilde{\beta} = 0.00017(P = 0.964)$

第2段階の推定式にOLSを用いると

$\hat{\beta} = 0.003(P = 0.009)$

推定値の偏りが調整された結果、
IQは賃金に影響を与えないことが示された ???

→ 検定

内生性の検証

(4) Durbin-Wu-Hausman検定

```
. estat endog
```

```
Tests of endogeneity  
Ho: variables are exogenous  
  
Durbin (score) chi2(1)          = .457658    (p = 0.4987)  
Wu-Hausman F(1,744)            = .449477    (p = 0.5028)
```

結果がメイン画面に出力

Command

estat endog ← “estat endog”と入力

ここでは、帰無仮説（第1段階の誤差項は目的変数と独立）が棄却される → *IQscore*は内生変数と考えられる

過剰識別制約の検定を実行

(4) 過剰識別制約の検定

```
. estat overid
```

```
Tests of overidentifying restrictions:
```

```
Sargan (score) chi2(3) = 87.6552 (p = 0.0000)  
Basmann chi2(3)      = 97.025 (p = 0.0000)
```

結果がメイン画面に出力

Command

estat overid ← “estat overid”と入力

ここでは、帰無仮説（操作変数と誤差項は無相関）が棄却される
→ 操作変数の選択に問題がある

弱い操作変数かどうかを検証

(4)弱い操作変数
かどうか検証

```
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,742)	Prob > F
iq	0.3360	0.3226	0.0692	13.7859	0.0000

$F = 13.8$

結果が
メイン
画面に
出力

Minimum eigenvalue statistic = 13.7859

Critical values

Ho: Instruments are weak

of endogenous regressors:

1

of excluded instruments:

4

2SLS relative bias

5%	10%	20%	30%
16.85	10.27	6.71	5.34

2SLS Size of nominal 5% wald test

10%	15%	20%	25%
24.58	13.96	10.26	8.31

LIML Size of nominal 5% wald test

5.44	3.87	3.30	2.98
------	------	------	------

“estat firststage”

と入力 →

Command

```
estat firststage
```

$F > 10 \rightarrow$ 操作変数は弱くないように見える

弱い操作変数かどうかを検証

```
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,742)	Prob > F
iq	0.3360	0.3226	0.0692	13.7859	0.0000

$F = 13.8$ ←

Minimum eigenvalue statistic = 13.7859

Critical values

Ho: Instruments are weak

of endogenous regressors: 1

of excluded instruments: 4

	5%	10%	20%	30%
2SLS relative bias	16.85	10.27	6.71	5.34
2SLS Size of nominal 5% wald test	24.58	13.96	10.26	8.31
LIML Size of nominal 5% wald test	5.44	3.87	3.30	2.98

OLSのバイアスと比べて、
2SLSのバイアスは何%程度か

Command

```
estat firststage
```

$10.27 < F < 16.85 \rightarrow 5 \sim 10\% \text{ほどのバイアス} \rightarrow \text{許容?}$

弱い操作変数かどうかを検証

```
. estat firststage
```

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(4,742)	Prob > F
iq	0.3360	0.3226	0.0692	13.7859	0.0000

$F = 13.8$ ←

Minimum eigenvalue statistic = 13.7859

Critical values

Ho: Instruments are weak

of endogenous regressors: 1

of excluded instruments: 4

	5%	10%	20%	30%
2SLS relative bias	16.85	10.27	6.71	5.34
2SLS size of nominal 5% wald test	24.58	13.96	10.26	8.31
LIML size of nominal 5% wald test	3.44	3.87	3.30	2.98

実際には何%水準で

帰無仮説(弱い操作変数である)を棄却できるか

```
Command
```

```
estat firststage
```

$10.26 < F < 13.96 \rightarrow 20\%$ 水準で棄却 $\rightarrow$ 弱い? 弱くない?

過剰識別・弱い操作変数の追加検証

(5)追加検証

結果が
メイン
画面に
出力

weak instrument robust tests for linear IV
H0: $\beta_{lw:iq} = 0$

Test	Statistic		p-value	
CLR	stat(.)	= 43.36	Prob > stat =	0.0000
AR	chi2(4)	= 98.99	Prob > chi2 =	0.0000
LM	chi2(1)	= 1.76	Prob > chi2 =	0.1841
J	chi2(3)	= 97.23	Prob > chi2 =	0.0000
LM-J	H0 rejected at 5% level			
wald	chi2(1)	= 0.00	Prob > chi2 =	0.9643

Note: wald test not robust to weak instruments.

Command

“rivtest”と入力 → rivtest

※Stataに搭載されていない複数の検定を実行できる

※内生変数を持つプロビットモデル・トービットモデルにも適用可能

※不均一分散を考慮した標準誤差を適用した場合も検定可能

過剰識別・弱い操作変数の追加検証

条件付き尤度比検定
(CLR検定)

帰無仮説
(弱い操作変数である)
に対して $P = 0.000$

weak instrument robust tests for linear IV
 $H_0: \text{beta}[\text{lw:iq}] = 0$

Test	Statistic		p-value	
CLR	stat(.)	= 43.36	Prob > stat =	0.0000
AR	chi2(4)	= 98.99	Prob > chi2 =	0.0000
LM	chi2(1)	= 1.76	Prob > chi2 =	0.1841
J	chi2(3)	= 97.23	Prob > chi2 =	0.0000
LM-J	H0 rejected at 5% level			
wald	chi2(1)	= 0.00	Prob > chi2 =	0.9643

Note: wald test not robust to weak instruments.

Command

`rivtest`

過剰識別制約の検定

帰無仮説
(操作変数と誤差項
とは無相関)
に対して $P = 0.000$

弱い操作変数との帰無仮説は棄却されるが、
操作変数と誤差項との相関が疑われる

(問) 操作変数の選択

推定式

$$\log(wage) = X\alpha + \beta IQscore + e$$

$$IQscore = X\gamma + Z\theta + u$$

Z: 操作変数(の候補)

med(母親の教育年数)

kww(別の共通試験(Knowledge in World of Workテスト)のスコア)

age(年齢)

mrt(結婚している=1)

(問)

操作変数の候補(*med*, *kww*, *age*, *mrt*)をすべて投入すると、過剰識別制約の検定をパスできません。どの変数を組み合わせれば検定をパスできるでしょうか？

うまくいった場合の結果

第2段階の推定結果

Instrumental variables (2SLS) regression

Number of obs = 758
wald chi2(12) = 372.05
Prob > chi2 = 0.0000
R-squared = 0.0944
Root MSE = .40793

lw	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
iq	.0243202	.0059992	4.05	0.000	.012562	.0360783
s	.0004625	.0189937	0.02	0.981	-.0367644	.0376893
expr	.039129	.0084408	4.64	0.000	.0225854	.0556726
tenure	.0327048	.009695	3.37	0.001	.0137029	.0517067
rns	-.0341617	.0383277	-0.89	0.373	-.1092825	.0409592
smsa	.1140326	.0336048	3.39	0.001	.0481683	.1798968
_Iyear_67	-.0679321	.0599189	-1.13	0.257	-.185371	.0495068
_Iyear_68	.0900522	.0561664	1.60	0.109	-.0200319	.2001363
_Iyear_69	.1794505	.0553432	3.24	0.001	.0709799	.2879212
_Iyear_70	.1395755	.0655531	2.13	0.033	.0110938	.2680573
_Iyear_71	.1735613	.0554814	3.13	0.002	.0648198	.2823028
_Iyear_73	.2971599	.0512878	5.79	0.000	.1966376	.3976821
_cons	2.837153	.4047262	7.01	0.000	2.043904	3.630402

*IQscore*の推定値は0.024($P = 0.000$)

うまくいった場合の結果

検定結果

Tests of overidentifying restrictions:

Sargan (score) $\chi^2(1) = .657488$ (p = **0.4174**)
 Basman $\chi^2(1) = .645904$ (p = **0.4216**)

estat firststage

First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	F(2,744)	Prob > F
iq	0.3196	0.3077	0.0462	18.0051	0.0000

weak instrument robust tests for linear IV
H0: beta[lw:iq] = 0

Test	Statistic		p-value	
CLR	stat(.)	= 27.25	Prob > stat	= 0.0000
AR	$\chi^2(2)$	= 27.90	Prob > χ^2	= 0.0000
LM	$\chi^2(1)$	= 26.71	Prob > χ^2	= 0.0000
J	$\chi^2(1)$	= 1.19	Prob > χ^2	= 0.2754
LM-J	H0 rejected at 5% level			
Wald	$\chi^2(1)$	= 16.43	Prob > χ^2	= 0.0001

Minimum eigenvalue statistic = **18.0051**

Critical values # of endogenous regressors: **1**
 Ho: Instruments are weak # of excluded instruments: **2**

	5%	10%	20%	30%
2SLS relative bias		(not available)		
2SLS Size of nominal 5% wald test	10% 19.93	15% 11.59	20% 8.75	25% 7.25
LIML Size of nominal 5% wald test	8.68	5.33	4.42	3.92

過剰識別制約の検定はOK

弱い操作変数の心配も少ないと考えられる

(参考) OLSの結果

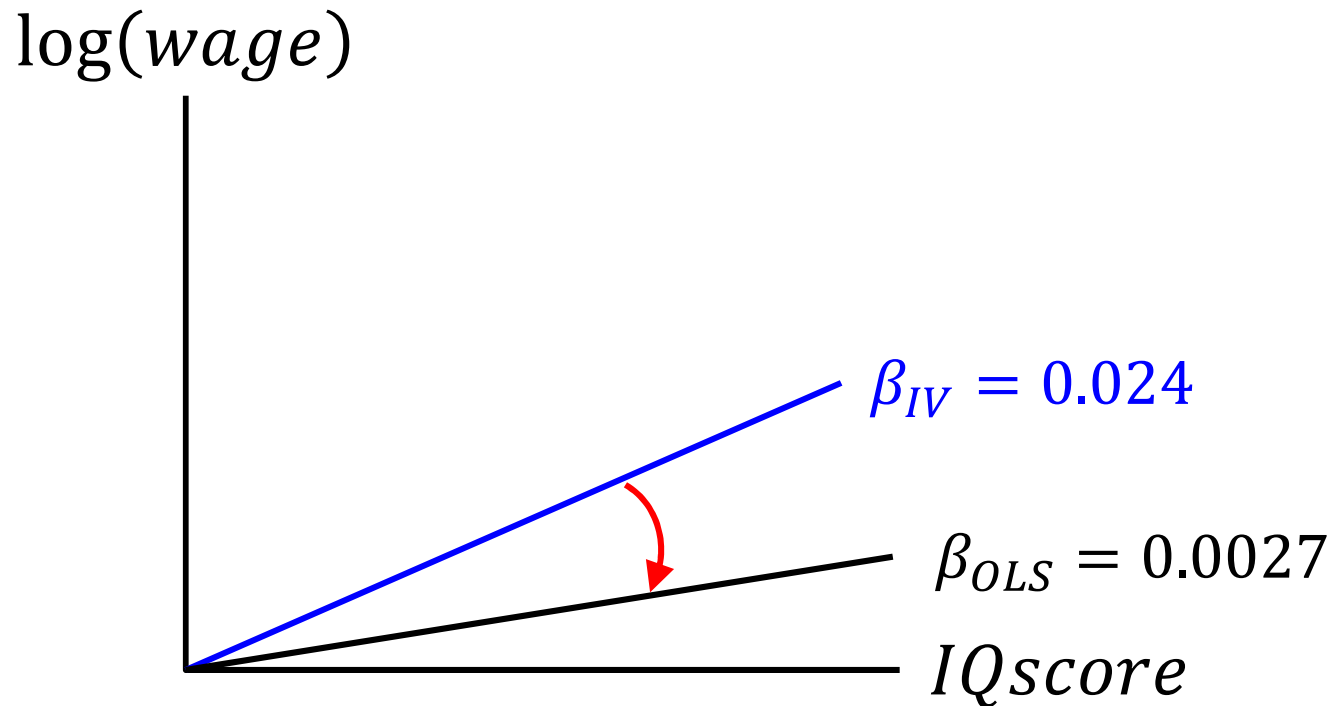
第2段階の推定結果

Source	SS	df	MS	Number of obs = 758		
Model	59.9127611	12	4.99273009	F(12, 745) = 46.86		
Residual	79.3733888	745	.106541461	Prob > F = 0.0000		
Total	139.28615	757	.183997556	R-squared = 0.4301		
				Adj R-squared = 0.4210		
				Root MSE = .32641		

lw	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
iq	.0027121	.0010314	2.63	0.009	.0006873	.0047369
s	.0619548	.0072786	8.51	0.000	.0476658	.0762438
expr	.0308395	.0065101	4.74	0.000	.0180592	.0436198
tenure	.0421631	.0074812	5.64	0.000	.0274763	.0568498
rns	-.0962935	.0275467	-3.50	0.001	-.1503719	-.0422151
smsa	.1328993	.0265758	5.00	0.000	.0807268	.1850717
_Iyear_67	-.0542095	.0478522	-1.13	0.258	-.1481506	.0397317
_Iyear_68	.0805808	.0448951	1.79	0.073	-.0075551	.1687168
_Iyear_69	.2075915	.0438605	4.73	0.000	.1214867	.2936963
_Iyear_70	.2282237	.0487994	4.68	0.000	.132423	.3240245
_Iyear_71	.2226915	.0430952	5.17	0.000	.1380889	.307294
_Iyear_73	.3228747	.0406574	7.94	0.000	.2430579	.4026915
_cons	4.235357	.1133489	37.37	0.000	4.012836	4.457878

*IQscore*の推定値は0.0027($P = 0.009$)

推定値の偏り(OLSと2SLS)



この例では、
OLSによる結果は過小推計と考えられる

(参考) 丁度識別と過剰識別

$$\log(wage) = X\alpha + \beta IQscore + e$$

$$IQscore = X\gamma + Z\theta + u$$

	丁度識別 (Just-identification)	過剰識別 (Over-identification)
Zの数	内生変数の数と同じ	内生変数の数より多い
Zと誤差項e との相関	検証しづらい	過剰識別制約の検定 ("estat overid"など)
弱い 操作変数	検証可能 ("estat firststage"など)	
目的変数の 内生性	検証可能 ("estat endog"など)	

【パート3】データを用いた実習(2)

2. 内生変数を持つプロビットモデル(IVP)

Health and Retirement Study (2002, wave 5)

(出典) Cameron, A. C. and P. K. Trivedi (2009) *Microeconometrics Using Stata*, Stata Press. 第14章を改変

データファイル: mus14data.dta

推定式: $retire^* = X\alpha + \beta \log(hincome) + e$ (N=3,197)

目的変数

(退職の有無)

$retire = 0$

$retire = 1$ (退職した)

内生変数(世帯所得(対数))

✓ 職場の労働環境

✓ 仕事をするのが嫌い

など、

未観測情報との相関が疑われる

【パート3】データを用いた実習

$\log(hincome)$ を操作変数に回帰するための式を追加

$$\log(hincome) = X\gamma + \textcolor{red}{Z}\theta + u$$

X : 独立変数(外生)

female(女性=1)

age(年齢)

age2(= age^2)

educyear(教育年数)

married(既婚=1)

chronic(慢性疾患数, 0~8)

Z : 操作変数

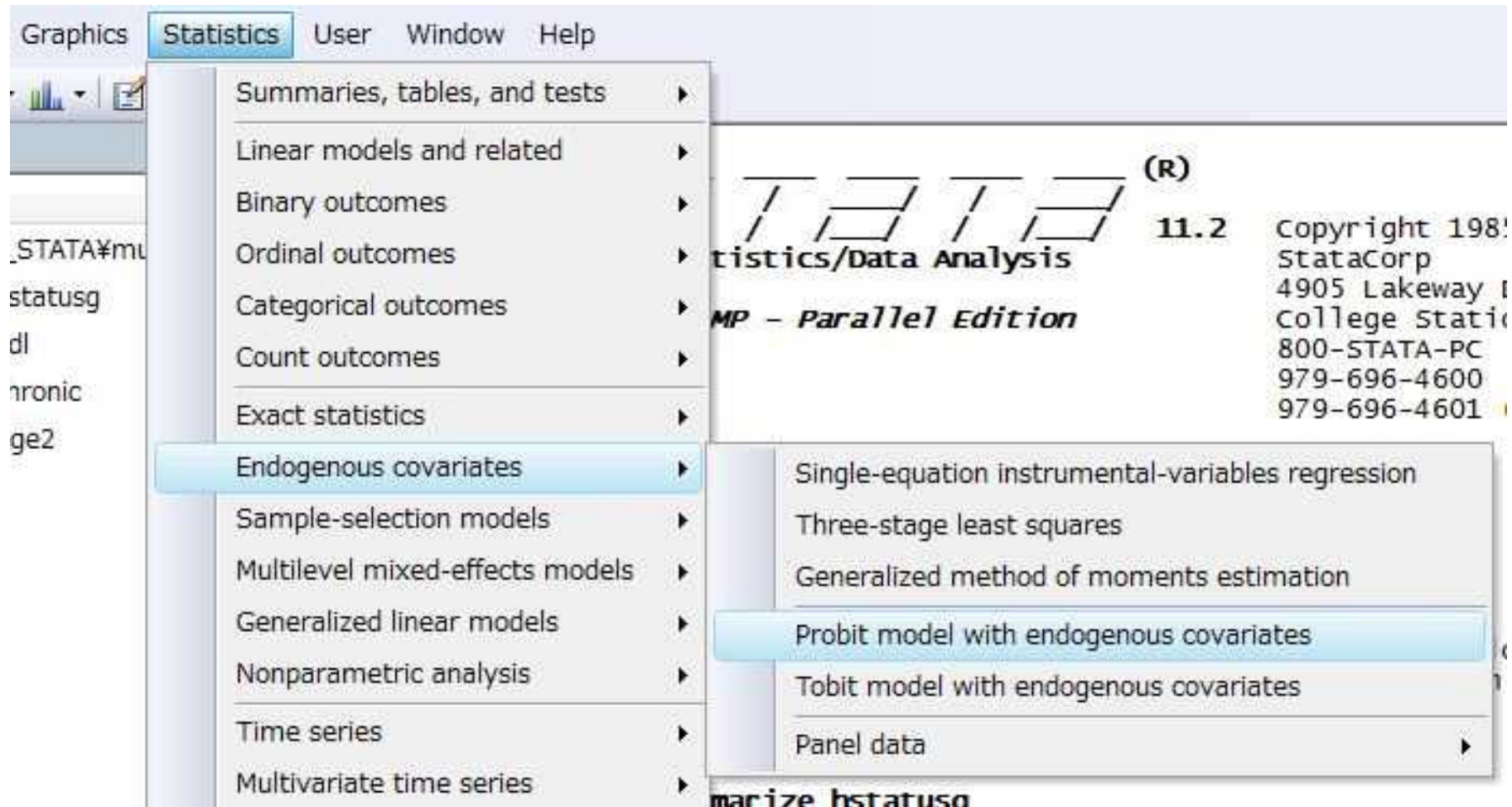
ins(私的保険加入=1)

hisp(ヒスパニック=1)

→ IVPをStataで実行しましょう

StataでIVPを実行するには

(1) Stataのツールバーから、Statistics → Endogenous covariates
→ “Probit model with endogenous covariates”を選択



StataでIVPを実行するには

(2)項目を選択or 入力

独立変数(群)

目的変数

内生変数

操作変数(群)

※過剰識別

推定方法

※ここでは

Two-step estimator

The screenshot shows the 'ivprobit' dialog box in Stata. The 'Model' tab is selected. The 'Dependent variable' is 'retire'. The 'Independent variables' are 'female age age2 educyear married chronic'. The 'Endogenous variables' are 'linc'. The 'Instrumental variables' are 'ins hisp'. The 'Use Newey's two-step estimator instead of maximum likelihood estimator' checkbox is checked. The 'Retain perfect predictor variables' checkbox is unchecked. The 'Constraints: (optional)' field is empty. The 'OK', 'Cancel', and 'Submit' buttons are at the bottom.

Annotations with arrows pointing to the dialog box:

- 目的変数 (Dependent variable)
- 内生変数 (Endogenous variables)
- 操作変数(群) (Instrumental variables)
- ※過剰識別 (Overidentification)
- 推定方法 (Estimation method)
- ※ここでは Two-step estimator (Here, Two-step estimator)

※Reportingのタブで

“Report first-stage regression”を選択すると
第1段階・第2段階両方の結果を出力

StataでIVPを実行するには

(3)結果がメイン画面に出力(第1段階の推定結果)

First-stage regression

Source	SS	df	MS
Model	1168.60968	8	146.07621
Residual	1651.54911	3188	.518051792
Total	2820.15879	3196	.882402626

Number of obs = 3197
F(8, 3188) = 281.97
Prob > F = 0.0000
R-squared = 0.4144
Adj R-squared = 0.4129
Root MSE = .71976

linc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ins	.2947565	.0271828	10.84	0.000	.2414589	.3480541
hisp	-.1915395	.0521061	-3.68	0.000	-.2937044	-.0893746
female	-.078106	.0277807	-2.81	0.005	-.1325758	-.0236361
age	.3134189	.0613745	5.11	0.000	.1930814	.4337564
age2	-.002234	.000456	-4.90	0.000	-.0031281	-.00134
educyear	.0965574	.0042392	22.78	0.000	.0882456	.1048691
married	.7720169	.0310242	24.88	0.000	.7111876	.8328463
chronic	-.0716464	.0092969	-7.71	0.000	-.0898749	-.053418
_cons	-9.188245	2.066824	-4.45	0.000	-13.24068	-5.135806

内生変数を操作変数に回帰した結果

StataでIVPを実行するには

(3)結果がメイン画面に出力(第2段階の推定結果)

Two-step probit with endogenous regressors

Number of obs = 3197
wald chi2(7) = 426.69
Prob > chi2 = 0.0000

内生変数

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
→ linc	.3235084	.1653558	1.96	0.050	-.000583	.6475998
female	-.503823	.0547636	-9.20	0.000	-.6111577	-.3964883
age	.8302306	.1560108	5.32	0.000	.5244552	1.136006
age2	-.0054031	.0011501	-4.70	0.000	-.0076572	-.003149
educyear	.0110557	.0199548	0.55	0.580	-.0280549	.0501664
married	-.2260512	.1461793	-1.55	0.122	-.5125573	.0604549
chronic	.033284	.0212281	1.57	0.117	-.0083222	.0748902
_cons	-31.83139	5.192192	-6.13	0.000	-42.0079	-21.65488

Instrumented: linc

Instruments: female age age2 educyear married chronic ins hisp

wald test of exogeneity: chi2(1) = 7.27 Prob > chi2 = 0.0070

世帯所得は退職決定に影響する($\beta_{linc} = 0.324$)

内生性の検証

(4) 外生性の検定結果を確認

Two-step probit with endogenous regressors

Number of obs = 3197
wald chi2(7) = 426.69
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
linc	.3235084	.1653558	1.96	0.050	-.000583	.6475998
female	-.503823	.0547636	-9.20	0.000	-.6111577	-.3964883
age	.8302306	.1560108	5.32	0.000	.5244552	1.136006
age2	-.0054031	.0011501	-4.70	0.000	-.0076572	-.003149
educyear	.0110557	.0199548	0.55	0.580	-.0280549	.0501664
married	-.2260512	.1461793	-1.55	0.122	-.5125573	.0604549
chronic	.033284	.0212281	1.57	0.117	-.0083222	.0748902
_cons	-31.83139	5.192192	-6.13	0.000	-42.0079	-21.65488

Instrumented: linc

Instruments: female age age2 educyear married chronic ins hisp

Wald test of exogeneity: chi2(1) = 7.27 Prob > chi2 = 0.0070

帰無仮説: 2本の推定式の誤差項は独立
→ 帰無仮説は棄却

(補足) estat ***コマンドとIVP

”estat***”を実行すると・・・

```
. estat endog
invalid subcommand endog
r(321);

. estat overid
invalid subcommand overid
r(321);

. estat firststage
invalid subcommand firststage
r(321);
```

“estat endog”, “estat overid”, “estat firststage”の各コマンドは2SLS(とGMM)にしか適用できない(みたいです)

OLSとの比較

$$\beta_{OLS} = -0.102 (P = 0.002)$$



符号が変化

$$\beta_{IV} = 0.324 (P = 0.050)$$

この例では、
OLSによる結果は過小推計と考えられる
世帯所得は退職決定に正の影響を与える

補足(メモ)

(1)誤差項の問題

誤差項が*i.i.d.*に従うもとの推定 → 2SLSを使用

誤差項に系列相関など、複雑な構造がある場合

→ 一般化モーメント法(GMM)を使用(より弱い制約を適用)

誤差項に不均一分散が確認できる場合

→ 不均一分散に対してロバストな標準誤差を適用

Stataでは、コマンドラインに`vce(robust)`等と書き加えることで実行可

(2)操作変数法的前提

理論的に、大標本のもとでバイアス修正が可能な推定方法

→ 小標本では、かえってバイアスが大きくなることに注意

IVに関する参考文献

[1] Wooldridge, J. (2012) *Introductory Econometrics* (5th ed.), Cengage Learning.

※第15章に2SLSとその検定の考え方が解説

[2] Stock, J. H. and M. W. Watson (2010) *Introduction to Econometrics* (3rd ed.), Addison-Wesley.

※著者の一人は操作変数法の発展に寄与

[3] Angrist, J. D. and J.-S. Pischke (2009) *Mostly Harmless Econometrics*, Princeton University Press.

(日本語訳)大森義明・小原美紀・田中隆一・野口晴子(訳)『「ほとんど無害」な計量経済学』、NTT出版。

※操作変数法に関する最新の流れを解説

[4] Cameron, A. C. and P. K. Trivedi (2010) *Microeconometrics Using Stata* (rev. ed.), Stata Press.

※IVPに関する記述あり

[5] 北村行伸(2009)『ミクロ計量経済学入門』、日本評論社。

検定方法に関する参考文献

（弱い操作変数の検証）

- [1] Staiger, D. and J. H. Stock (1997) “Instrumental Variables Regression with Weak Instruments”, *Econometrica*, 65(3), 557-586.
- [2] Stock, J. H., J. H. Wright and M. Yogo (2002) “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments”, *Journal of Business and Economic Statistics*, 20, 518-529.

（“rivtest”コマンド）

- [3] Finlay, K. and L. M. Magnusson (2009) “Implementing Weak-Instrument Robust Tests for a General Class of Instrumental-Variables Models”, *Stata Journal*, 9(3), 1-24.

（CLR検定）

- [4] Moreira, M. J. (2003) “A Conditional Likelihood Ratio Test for Structural Models”, *Econometrica*, 71(4), 1027-1048.
- [5] Magnusson, L. M. (2010) “Inference in Limited Dependent Variable Models Robust to Weak Identification”, *Econometrics Journal*, 13(3), S56-S79.