# 第 12 章 操作変数回帰分析

Stock & Watson 著、宮尾訳 (2016)『入門計量経済学』

資料作成：田中 鮎夢

2025-02-09

# はじめに

第９章：誤差項が説明変数と相関を持つという問題を議論

1. 除外された変数
2. 変数の計測誤差
3. 同時双方向の因果関係（因果関係が $X$ から $Y$、そして $Y$ から $X$ へと両方向に存在する）

除外された変数のバイアスについては、多変数回帰モデルに除外された変数自身を含めることで直接対処できるが、それはそのデータがある場合にのみ可能な方法。

同時双方向の因果関係のように、直接の対処方法がそもそも存在しない、または利用できないとき、新しい方法が必要。

# 操作変数回帰

**操作変数回帰 (instrumental variables 〈IV〉 regression)**

説明変数 $X$ が誤差項 $u$ と相関を持つときに、回帰係数の一致推定量を得るための一般的な方法。

$X$ の変動を 2 つの部分に分けて考える。

1. 何らかの理由で $u$ と相関がある部分 (この部分がバイアスを引き起こす)
2. $u$ とは相関しない部分

# IV 回帰が行うこと

**IV 回帰が行うこと**

何らかの情報を使って 2 つ目の部分を切り離すことができるなら、

$u$ と相関のない $X$ の変動だけに焦点を絞ることができ、

バイアスをもたらす $X$ の変動については考えずに済む。

# 操作変数

$u$ と相関のない $X$ の変動に関する情報は、操作変数 (instrumental variables あるいは instruments) と 呼ばれる追加的な変数から集める。

操作変数回帰は、これらの追加的な変数を「道具 (instrument)」として利用。

$\longrightarrow$ 無相関な $X$ の変動部分を取り出し、そこから回帰係数の一致推定量を求める。

# 本章の構成

## 12.1 節・12.2 節　IV 回帰の仕組みと仮定

- IV 回帰分析はなぜうまく機能するのか。
- 適切な操作変数とは何か。
- 最も典型的な IV 回帰の手法である 2 段階最小二乗法の使い方と解釈。

## 12.3 節 操作変数が正当かどうかの評価方法の問題

## 12.4 節 具体例「たばこ需要の弾力性を IV 回帰により推定」

## 12.5 節 正当な操作変数をどう探すか

# 12.1 操作変数法による推定：1 説明変数、1 操作変数の場合

設定

- 説明変数 $X$ が 1 つ
- 説明変数 $X$ が誤差項 $u$ との間に相関があるかもしれない

もし $X$ と $u$ が相関すれば、OLS 推定量は一致性を持たない。

$\longrightarrow$ 標本数が非常に多くても推定量は回帰係数の真の値には近づかない。((6.1) 式参照)

$\longrightarrow$ 相関の理由を問わず、適切な操作変数 $Z$ があれば、$X$ の 1 単位変化の $Y$ への影響は操作変数推定量を使って推定できる。

# 操作変数モデルとその仮定

$Y_i$ を $X_i$ で説明する母集団回帰式

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \ldots, n. \tag{12.1}$$

- $u_i$ は誤差項。$Y_i$ を決定する他のさまざまな要因を含む。

もし $X_i$ と $u_i$ に相関があれば、OLS 推定量は一致性を持たない。

操作変数法推定では、操作変数 $Z$ を使って，$u_i$ とは無相関な $X$ の変動分を取り出し、推定に利用。

# 内生性と外生性

操作変数回帰には、誤差項と相関がある変数と相関がない変数を区別するための特別の用語があります.

**内生変数 (endogenous variable)**: 誤差項と相関がある変数（モデル内部で決定される変数）

**外生変数 (exogenous variable)**: 誤差項と相関がない変数（モデル外部で決定される変数）

# 内生性と外生性の例

**2 つの因果関係の可能性**

生徒・教師比率高い $\longrightarrow$ テスト成績高い

テスト成績低い $\longrightarrow$ 政治的な介入や財政資金投入（削減）により生徒・教師比率低下

この例では、生徒・教師比率もテスト成績も内生的な変数。$\Rightarrow$ どちらも母集団の誤差項と相関

# 操作変数が正当であるための2つの条件

正当な操作変数は2つの条件を満たす必要がある。

1. 操作変数の妥当性 (instrument relevance)：$\mathrm{corr}\left(Z_i, X_i\right) \neq 0$.
   - 操作変数が妥当ならば、操作変数の変動は説明変数 $X_i$ の変動と関連。
2. 操作変数の外生性 (instrument exogeneity)：$\mathrm{corr}\left(Z_i, u_i\right) = 0$.
   - 操作変数が外生的であれば、操作変数によって捉えられる $X_i$ の変動も外生的。

$\longrightarrow$ 妥当で外生的な操作変数は、$X_i$ の外生的な変動を捉えることができる。

$\longrightarrow$ 外生的な変動が母集団の（つまり真の）係数 $\beta_1$ の推定に使われる。

# 2 段階最小二乗法

もし操作変数 Z が妥当性と外生性の条件を満たすならば、係数 $\beta_1$ は、2 段階最小二乗法 (two stage least squares: TSLS) と呼ばれる IV 推定量を使って推定できる。

その名前が表すように、この推定方法は 2 段階で計算される。

第 1 段階では X が 2 つの部分に分けられる：

1. 誤差項と相関する問題の部分
2. 誤差項と相関のない (つまり問題のない) 部分

第 2 段階では、問題のない部分を使って $\beta_1$ を推定。

# 第 1 段階

$X$ と $Z$ の関係に関する母集団回帰式。

$$X_i = \pi_0 + \pi_1 Z_i + v_i.$$

- $\pi_0$ は切片
- $\pi_1$ は傾き
- $v_i$ は誤差項

# 分解

この回帰式により $X_i$ の変動が 2 つに分解される。

1. $\pi_0 + \pi_1 Z_i$
   - $X_i$ の変動のうち $Z_i$ によって予測される部分
   - $Z_i$ は外生的なので、この部分の $X_i$ の変動は (12.1) 式の誤差項 $u_i$ とは相関を持たない
2. $v_i$
   - $u_i$ と相関を持つ問題の部分。

**2 段階最小二乗法の背後にある考え方**

- $X_i$ 変動の問題ない部分である $\pi_0 + \pi_1 Z_i$ を利用。
- $v_i$ の方は利用せず捨ててしまう。

TSLS の第 1 段階は (12.2) 式に OLS を当てはめ、OLS 回帰からの予測値

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

を利用。

($\hat{\pi}_0$ と $\hat{\pi}_0$ は OLS 推定値)

# 第 2 段階

**TSLS の第 2 段階**

$Y_i$ を $\hat{X}_i$ で OLS 回帰。

$\longrightarrow$ この 2 段階目の回帰から得られた推定量が TSLS 推定量 $\hat{\beta}_0^{\mathsf{TSLS}}$, $\hat{\beta}_1^{\mathsf{TSLS}}$

# 具体例 1 フィリップ・ライトの問題

フィリップ・ライトの 1928 年の著作『動物性油と植物性油への関税』の付論において、操作変数法が初めて公表された。

この付論は、彼の息子であり著名な統計学者であるセウォール・ライト (Sewall Wright) が 書いたという説もある。

動物性や植物性の油や脂肪 (バターや大豆油など) に対する輸入関税 (輸入品に対する税金) について、関税の経済効果を理解したい。

$\longrightarrow$ その財の需要、供給曲線の形状について具体的な数値が必要。

$\longrightarrow$ 供給の価格弾力性（価格の 1%上昇によって生じる供給量のパーセント変化)、需要の価格弾力性（価格の 1%上昇による需要量のパーセント変化）の推定値が必要。

# バター需要の式

バター需要の式

$$\ln\left(Q_i^{\mathsf{butter}}\right) = \beta_0 + \beta_1 \ln\left(P_i^{\mathsf{butter}}\right) + u_i. \tag{12.3}$$

- $Q_i^{\mathsf{butter}}$ はバター消費量の第 $i$ 観測値
- $P_i^{\mathsf{butter}}$ はバター価格の第 $i$ 観測値
- $u_i$ はバター需要に影響するその他の要因（所得や消費者の好みなど）を反映した誤差項

# バターの需要に関する弾力性

基本概念 8.2 より，$\ln(Y_i)$ を $\ln(X_i)$ で説明する式の係数は、$X$ に関する $Y$ の弾力性と解釈される。

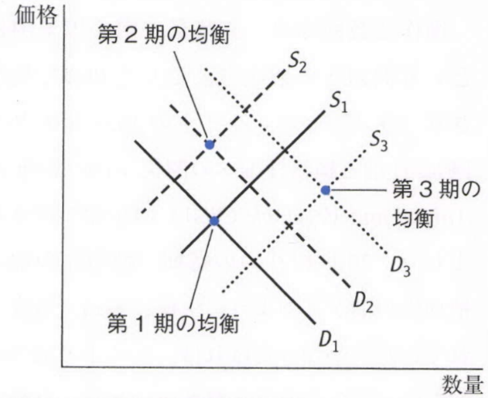
バター価格の 1% 上昇は $\beta_1$% の需要の変化をもたらす（$\beta_1$ が需要の弾力性)。

**需要と供給の相互関係から、説明変数 $\ln P_i^{\textbf{butter}}$ は 誤差項 $u_i$ と相関がある可能性。**

# 図 12.1 バターの需要曲線

## （a）均衡

第 1 期の均衡は需要曲線 $D_1$ と供給曲線 $S_1$ の 交点
第 2 期の均衡は $D_2$ と $S_2$ の交点
第 3 期の均衡は $D_3$ と $S_3$ の交点

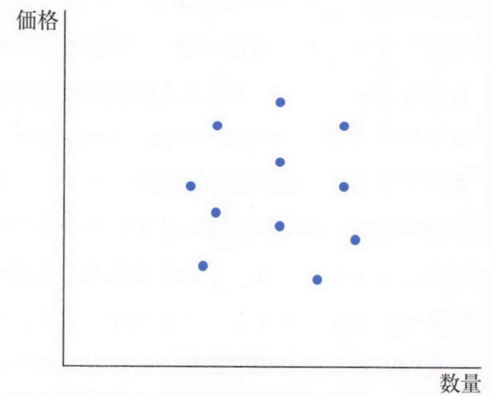


**(a)** 3 つの時点での需要と供給

# 図 12.1 バターの需要曲線

## (b) 散布図

11 時点の均衡価格と数量価格から、需要曲線と供給曲線を区別し決定することはできない。

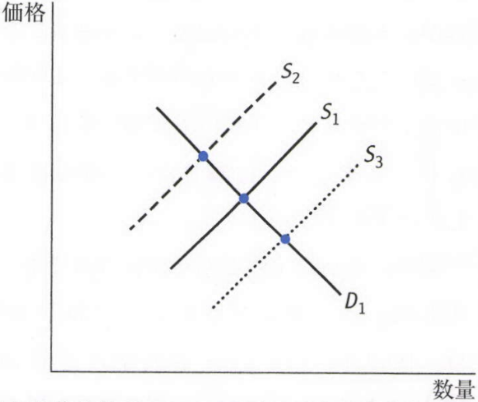



**(b)** 11 の時点での均衡価格と数量

# 図 12.1 バターの需要曲線

## （c）供給曲線のシフト

天候などの外生要因で供給曲線のみが $S_1, S_2, S_3$ とシフト
しかし需要曲線は $D_1$ のままであるとき
均衡価格と数量は需要曲線に沿って移動




**(c)** 供給曲線だけがシフトするときの均衡価格と数量

# 第 3 の変数

**操作変数**

ライトは、問題を回避する方法は、供給だけを変化させ、需要には影響を及ぼさないような第 3 の変数を見つけることではないかと気づいた。

ライトの問題では、この第 3 の変数が操作変数。

1. 価格と相関がある (それは供給を変化させ価格変化をもたらす)
2. しかし $u$ とは相関がない (需要関数は安定したまま)

# 潜在的な操作変数

操作変数候補：天候（酪農地域の降雨量）

## 妥当性の条件

酪農地域における降雨量が例年の平均よりも少なければ、牧草量が減少

$\rightarrow$ ある価格水準の下での生産量は減少

$\rightarrow$ 供給曲線が左へシフト、均衡価格は上昇（妥当性）

## 外生性の条件

一方、酪農地域の降雨量はバターの需要へ直接の影響を持つとは考えられない

$\rightarrow$ 降雨量と $u$ との相関はゼロ（外生性）

# 具体例 2: クラス人数がテスト成績へ及ぼす効果

第 II 部では、クラス人数がテスト成績へ及ぼす効果を推定

その際、生徒の特性や地区の特徴をコントロール

本章では、残った除外された変数のバイアス（校外での学習機会や教師の質など）を回避するために、操作変数法を利用

# 仮想的な例

カリフォルニア州で地震が発生

⟶ いくつかの学校が校舎の修理のために閉鎖（震源地により近い学区ほど損害は大きい）

⟶ 閉鎖校を含む学区では、１校当たりの生徒数を倍にして、一時的にクラス人数を増やさなければならない

⟶ 震源地からの距離がクラス人数と相関（**妥当性の条件**）

震源地からの距離（**外生変数**）は、生徒のテスト成績に影響を及ぼす他の要因（生徒がまだ英語を学習しているかどうかなど）とは無関係（**外生性の条件**）

⟶ 震源地からの距離という操作変数は、除外された変数バイアスを回避してクラス人数のテスト成績への影響を推定するのに有用。

# TSLS 推定量の標本分布

TSLS 推定量の大標本での分布は単純

- TSLS 推定量は一致性を持つ。
- TSLS 推定量は正規分布に従う。

# TSLS 推定量の公式

説明変数 $X$ が 1 つ、操作変数 $Z$ が 1 つの場合の TSLS 推定量の公式

- $s_{ZY}$ を $Z$ と $Y$ との標本共分散
- $s_{ZX}$ を $Z$ と $X$ との標本共分散

付論 12.2 で示すように、1 つの操作変数の場合の TSLS 推定量：

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \tag{12.4}$$

# 標本数が大きいときの $\hat{\beta}_1^{TSLS}$ の標本分布

（12.4）式の公式は，$\hat{\beta}_1^{TSLS}$ が一致性を持ち、大標本の下で正規分布に従うという性質を示すのに使われる。

以下では議論の要約を説明。

数学的な詳細については付論 12.3 を参照。

$\hat{\beta}_1^{TSLS}$ の一致性は、

1. 操作変数 $Z_i$ が妥当で外生的という仮定
2. 標本共分散が一致性を持つ（真の共分散に収束する）という性質

とを組み合わせることで示すことができる。

# TSLS 推定量の一致性の証明

（12.1）式の $Y_i = \beta_0 + \beta_1 X_i + u_i$ から、以下の式を得る。

$$\begin{aligned}
&\mathrm{cov}\left(Z_i, Y_i\right) \\
&= \mathrm{cov}\left[Z_i, (\beta_0 + \beta X_i + u_i)\right] \\
&= \beta_1 \underbrace{\mathrm{cov}\left(Z_i, X_i\right)}_{\neq 0} + \underbrace{\mathrm{cov}\left(Z_i, u_i\right)}_{=0} \qquad (12.5)
\end{aligned}$$

- 2 つ目の等式の成立は共分散の性質（ (2.33) 式 ）から示すことができる。
- 操作変数の外生性の仮定から、$\mathrm{cov}\left(Z_i, u_i\right) = 0$
- 操作変数の妥当性の仮定から、$\mathrm{cov}\left(Z_i, X_i\right) \neq 0$。

# TSLS 推定量の一致性の証明

したがって、もし操作変数が正当なら、以下の式が成立する。

$$\beta_1 = \frac{\mathrm{cov}\left(Z_i, Y_i\right)}{\mathrm{cov}\left(Z_i, X_i\right)} = \frac{\text{Z と Y の真の共分散}}{\text{Z と X の真の共分散}} \tag{12.6}$$

3.7 節で議論したように、標本共分散は真の共分散の一致推定量。

したがって、

$$s_{ZY} \xrightarrow{p} \mathrm{cov}\left(Z_i, Y_i\right)$$

$$s_{ZX} \xrightarrow{p} \mathrm{cov}\left(Z_i, X_i\right)$$

# TSLS 推定量の一致性の証明

$\longrightarrow$（12.4）式と（12.6）式より、TSLS 推定量は一致性を持つ：

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} \xrightarrow{p} \frac{\mathrm{cov}\left(Z_i, Y_i\right)}{\mathrm{cov}\left(Z_i, X_i\right)} = \beta_1 \tag{12.7}$$

# TSLS 推定量の標本分布

（12.4）式の公式は、$\hat{\beta}_1^{TSLS}$ の標本分布が大標本で正規分布に従うという性質を示すのに も用いられる。

導出のロジックは、最小二乗推定量の場合と同じ。

TSLS 推定量は確率変数の平均で表される。

標本数が大きい場合、**中心極限定理**から確率変数の平均は正規分布に従うことが示される。

# TSLS 推定量の標本分布

（12.4）式の $\hat{\beta}_1^{TSLS}$ の分子は以下の式で与えられる。

$$s_{ZY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)$$

つまり $\left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right)$ の平均である。

# TSLS 推定量の標本分布

付論 12.3 の数式展開から、この平均化によって中心極限定理が適用され、大標本において $\hat{\beta}_1^{TSLS}$ は

$$N\left(\beta_1, \sigma^2_{\hat{\beta}_1^{\mathsf{TSLS}}}\right)$$

という正規分布に近似的に従うことになる。

# TSLS 推定量の標本分布

そこで分散は、以下の式で与えられる。

$$\sigma^2_{\hat{\beta}_1^{\mathsf{TSLS}}} = \frac{1}{n}\frac{\mathrm{var}\left[\left(Z_i - \mu_{\mathrm{Z}}\right)u_i\right]}{\left[\mathrm{cov}\left(Z_i, X_i\right)\right]^2} \tag{12.8}$$

分散 $\sigma^2_{\hat{\beta}_1^{\mathsf{TSLS}}}$ は、（12.8）式に現れる分散と共分散項を推定することで推定できる。

$\sigma^2_{\hat{\beta}_1^{\mathsf{TSLS}}}$ の推定値の平方根は、IV 推定量の標準誤差。

この導出は計量経済ソフトの TSLS コマンドで自動的に計算される。

# 信頼区間と仮説検定

$\hat{\beta}_1^{TSLS}$ が大標本で正規分布に従うため、$\beta_1$ に関する仮説検定は $t$ 統計量を求めて実施される。

95% の大標本信頼区間は、

$$\hat{\beta}_1^{TSLS} \pm 1.96 SE\left(\hat{\beta}_1^{TSLS}\right)$$

で与えられる。

# たばこ需要への応用

人々のたばこ消費を著しく減らすには、正確にどの程度の課税が必要か?

例）たばこ消費を 20%減少させるには、課税後の販売価格をいくらに設定すればよいか?

$\longrightarrow$ たばこ需要の弾力性に依存

例）価格弾力性が-1 なら、消費量を 20%減らすには価格を 20%上昇させなければならない。

例）価格弾力性が-0.5 なら、消費量を 20%減らすには価格を 40%上昇させなければならない。

# たばこ需要の弾力性の推定

1985–1995 年における 48 州の年次データ。(データの詳細は付論 12.1 を参照)

たばこ需要の弾力性を TSLS に基づいて推定。

当面、すべての結果は、1995 年のクロスセクションの州データに基づく。

# 操作変数

操作変数 $SalesTax_i$：一般の売上税に占めるたばこ税の割合

たばこ消費量 $Q_i^{cigarette}$: 各州のたばこ販売数 (1 人当たりパック数)

たばこ価格 $P_i^{cigarette}$: すべての税金を含むたばこ 1 パックの平均価格

# 操作変数の正当性に関する２つの条件

**操作変数の正当性チェック**

統計ツールによる検討も大切だが、状況判断も重要な役割を果たす。

1. 操作変数の妥当性

たばこ税の増大はトータルの販売価格 $P_i^{cigarette}$ を 上昇させる。
$\longrightarrow$ たばこ税は妥当性の条件を満たす

# 操作変数の正当性に関する２つの条件

## 2. 操作変数の外生性

たばこ税が外生であるためには、需要式の誤差項と無相関であることが必要。
つまり、たばこ税は、販売価格を通じた間接的な経路を通じてのみ、たばこ需要に影響を与える。

当面この外生性はもっともらしいと考える。

- 一般の売上税の税率は州ごとに異なり、各州で公共支出を賄うための租税の組合せ—売上税・所得税・固定資産税そしてその他の税の組合せ—が異なる。
- これらの組合せがどう選択されるかは、それぞれの州の政治的な判断により、たばこ需要に関連した要因とは無関係。

# TSLS

## 第 1 段階の回帰式

$$\ln \widehat{P_i^{cigarette}} = 4.63 + 0.031 \times SalesTax_i, \qquad R^2 = 47\%$$

## 第 2 段階の回帰式

$$\ln \widehat{Q_i^{cigarette}} = 9.72 - 1.081 \times \ln \widehat{P_i^{cigarette}}$$

たばこ需要は弾力的：
1%の価格上昇は、たばこ消費を 1.08%減少させる
ただし、所得を制御する必要あり。

# 12.2 一般的な操作変数回帰モデル

一般的な操作変数 (IV) 回帰モデルには、4 つの変数が含まれる。

- 被説明変数 $Y$
- 誤差項と相関を持つかもしれない内生的な複数の説明変数 $X's$(例：たばこ価格)
- 誤差項とは相関のない追加的な複数の説明変数 $W's$—「付加された外生変数 (included exogenous variables)」
- 複数の操作変数 $Z's$

IV 回帰を行うためには、内生的な説明変数 ($X's$) と少なくとも同じ数の操作変数 ($Z's$) が必要。

# 操作変数の数と内生的な説明変数の数との関係

1. 操作変数の数 $=$ 内生的な説明変数の数

$\longrightarrow$ 過不足なく識別される (exactly identified)

2. 操作変数の数 $>$ 内生的な説明変数の数

$\longrightarrow$ 過剰に識別される (overidentified)

3. 操作変数の数 $<$ 内生的な説明変数の数

$\longrightarrow$ 過少に識別される (underidentified)

係数が IV 回帰によって推定されるためには、操作変数の数は内生的な説明変数の数以上でなければならない（1. か 2.）。

# 基本概念 12.1

## 一般的な操作変数（IV）回帰モデル

$$Y_i = \beta_0 + \underbrace{\beta_1 X_{1i} + \cdots + \beta_k X_{ki}}_{\text{endogenous}} + \underbrace{\beta_{k+1} W_{1i} + \cdots + \beta_{k+r} W_{ri}}_{\text{exogenous}} + u_i$$

$$i = 1, \dots, n \tag{12.12}$$

- $Y_i$ は被説明変数
- $u_i$ は誤差項。変数の測定誤差および／あるいは除外された要因を反映。
- $X_{1i}, \dots, X_{ki}$ は $k$ 個の内生的な説明変数。潜在的に $u_i$ との相関を持つ。
- $W_{1i}, \dots, W_{ri}$ は $r$ 個の付加された外生的な説明変数。$u_i$ とは相関を持たない。
- $\beta_0, \beta_1, \dots, \beta_{k+r}$ は未知の回帰係数。
- $Z_{1i}, \dots, Z_{mi}$ は $m$ 個の操作変数。

# 一般的な操作変数モデルにおける TSLS

## 1つの内生的な説明変数を持つ2段階最小二乗法（TSLS）

内生的な説明変数が1つ、そして付加された外生変数がいくつかある場合の推計式：

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i \qquad (12.13)$$

ここで $X_i$ は誤差項と相関があるかもしれないが、$W_{1i}, \ldots, W_{ri}$ はそうではない。

# TSLS 回帰の第 1 段階

母集団における TSLS 回帰の第 1 段階は、$X$ を外生変数 $W'$s と操作変数 $Z'$s で回帰。

**X の誘導形 (reduced form) の式**

$$X_i = \pi_0 + \underbrace{\pi_1 Z_{1i} + \cdots + \pi_m Z_{mi}}_{\text{IV}} + \underbrace{\pi_{m+1} W_{1i} + \cdots + \pi_{m+r} W_{ri}}_{\text{exogenous}} + v_i. \tag{12.14}$$

$\pi_0, \pi_1, \ldots, \pi_{m+r}$ は未知の回帰係数

$v_i$ は誤差項

# TSLS 回帰の第 2 段階

第 1 段階の結果を使って、第 2 段階の TSLS 推定を行う。

**TSLS 推定量**

TSLS の第 2 段階では、(12.13) 式を OLS により推定。

ただし、第 1 段階で求められた $X$ の予測値を利用。

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + \beta_2 W_{1i} + \cdots + \beta_{1+r} W_{ri} + u_i \qquad (12.13')$$

# 基本概念 12.2

一般的な操作変数回帰モデル (12.12) 式における TSLS 推定量は 2 段階で計算される。

1. 第 1 段階の回帰（first-stage regression(s)）：

$X_{1i}$ を操作変数 $(Z_{1i}, \dots, Z_{mi})$ と付加された外生変数 $(W_{1i}, \dots, W_{ri})$ で OLS により回帰。

この回帰からの予測値を計算し、それを $\hat{X}_{1i}$ とする。

この推定をすべての内生的な説明変数 $X_{2i}, \dots, X_{ki}$ について繰り返し、予測値 $\hat{X}_{1i}, \dots, \hat{X}_{ki}$ を求める。

# 基本概念 12.2（続き）

2. 第 2 段階の回帰（second-stage regression(s)）：

$Y_i$ を、内生的な説明変数の予測値 $\left(\hat{X}_{1i}, \ldots, \hat{X}_{ki}\right)$ と付加された外生変数 $(W_{1i}, \ldots, W_{ri})$ で OLS により回帰。

TSLS 推定量 $\hat{\beta}_0^{TSLS}, \ldots, \hat{\beta}_{k+r}^{TSLS}$ は、この第 2 段階の回帰で得られた推定量。

# 複数の内生的な説明変数への拡張

内生的な説明変数が複数ある場合（$X_{1i}, \dots, X_{ki}$ の場合）でも、TSLS の計算方法は同様。

それぞれの内生変数について個別に第 1 段階の回帰が行われる：

- つまり被説明変数は複数ある $X's$ の中の 1 つ。
- 説明変数はすべての操作変数 ($Z's$) とすべての付加された外生変数 ($W's$)。

これらすべてを合わせて、第 1 段階の回帰によりそれぞれの内生的な説明変数の予測値が計算される。

TSLS の第 2 段階では、内生的な説明変数の予測値 ($X's$) を使って (12.12) 式が OLS を使って推定される。

# 基本概念 12.3 操作変数が正当であるための 2 条件

## 1. 操作変数の妥当性

$X$ が複数の場合、$X_{1i}$ を操作変数 $Z's$ と付加された外生変数 $W's$ で回帰。その予測値を $\hat{X}^*_{1i}$ とする。
すべての観測値が 1 を取る説明変数を「1」とする (その係数は定数項となる)。
このとき、$\left(\hat{X}^*_{1i'}, \ldots, \hat{X}^*_{ki'} W_{1i}, \ldots, W_{ri}, 1\right)$ に完全な多重共線性は存在しない。

$X$ が 1 つの場合、上記の条件が成立するためには、$X$ を説明する母集団の回帰式 (複数の $Z's$ と $W's$ が説明変数) に少なくとも 1 つの $Z$ が入っていなければならない。

# 基本概念 12.3 操作変数が正当であるための 2 条件

## 2. 操作変数の外生性

操作変数は誤差項と無相関である。

$$\mathrm{corr}\left(Z_{1i}, u_i\right) = 0, \ldots, \mathrm{corr}\left(Z_{mi}, u_i\right) = 0$$

# 操作変数回帰の仮定と TSLS の標本分布

IV 回帰の仮定の下、TSLS 推定量は一致性を持ち、その標本分布は大標本において正規分布に近似される。

操作変数回帰の仮定 IV 回帰の仮定は、多変数回帰モデルにおける最小二乗法の仮定 (基 本概念 6.4) を修正したもの