2021年10月5日

#### 計量経済学入門

西村　翼

# 第1回

### BIAS CAUSED BY OMISSION OF RELEVANT VARIABLES（説明変数の欠損によるバイアス）

回帰モデルの正しい仕様に基づく仮定に基づく分析は，以下のように表現される．

|  |  |
| --- | --- |
| ：被説明変数ベクトル（）  ：説明変数行列（）  ：（）ベクトル  ：誤差項ベクトル（） | (4.7) |

回帰モデルを構築する際に犯しうる仕様エラーには多くの種類がある．最も一般的なミスは，説明変数の省略と余分な（無関係な）変数を含めるしまうことである．

正しい仕様の回帰モデルは，次のように表現されると仮定する．

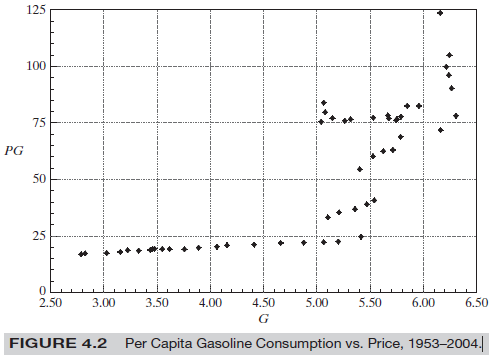
|  |  |
| --- | --- |
| ：説明変数行列（）  ：説明変数行列（）  ：（）ベクトル．  ：誤差項ベクトル（） | (4.8) |

を含まずにをに回帰すると，推定量は次の通りとなる．

|  |  |
| --- | --- |
|  | (4.9) |

(4.9)式の両辺に期待値をとることで， または でないならばバイアスが生じることが分かる．

|  |  |
| --- | --- |
|  | (4.10) |
| ただし，  ：（）行列であり，この行列の各列はの列にを最小二乗回帰した時の傾きの 　列である． | (4.11) |



Example 4.2　説明変数の欠落（Omitted Variable）

関連する所得変数なしで需要方程式を推定した場合，(4-10)は推定された価格弾力性にどのような偏りがあるかを示している．例2.3で検討したガソリン市場のデータが顕著な例となる．

を推定値とする．このとき， を所得係数とすると以下が得られる．

集計されたデータでは，欠損した共分散が正になるのか負になるのかは不明である．しかし，とはガソリンのような正常な財に対しては正であるため，の偏りの符号はこの共分散と同じになる．図4.2は，一人当たりのガソリン消費量G/Popを縦軸，価格指数PGを横軸として単純にプロットしたものです．このプロットは，人が期待するものとはかなり異なっている．しかし，付表F2.2のデータを見ると，何が起こっているのかがよく分かる．

一人当たりの所得や所得／人口，その他の物価を一定にすると，これらのデータは期待通りになるかもしれない．しかし，このデータでは所得が持続的に増加しており，G/Popと所得/Pop，PGと所得/Popの単純な相関は，それぞれ0.938，0.934とかなり大きい値を示している．価格と消費の間に期待される関係が現れるかどうかを確認するためには，データから所得／Popの介入効果を取り除く必要がある．

そのためには，定理3.2のFrisch-Waughの結果を利用する．一人当たりのガソリン消費量の対数を定数と物価指数の対数に回帰した単純回帰では，係数は0.29904で予想通り「間違った」符号を示した．一人当たりのガソリン消費量の対数を定数，物価指数の対数，一人当たりの所得の対数に重回帰すると，価格弾力性の推定値は-0.16949，所得弾力性の推定値は0.96595となる．

これは期待通りの結果である．この結果は，少なくともこの時期（1953年から2004年）の米国市場では，ガソリン消費量の変化の主な要因は価格の変化ではな，、所得（生産高）の伸びであるという，広く観察されている結果ともほぼ一致している．

このように考えると，説明変数が1つ，省略された変数が1つの場合の偏りの方向を推測するのは簡単である．しかし，2つ以上の説明変数が含まれている場合，省略変数式の項には多重回帰係数が含まれ，それ自体が単純相関ではなく偏相関の符号を持っていることに注意する必要がある．

例えば，先ほどの需要方程式では，もし関連性の高い製品の価格も含まれていたとすると，価格と所得の単純な相関だけでは，価格弾力性の偏りの方向性を判断するには不十分ある．求められるのは，価格と所得の相関から他の価格の影響を差し引いた値の符号である．この要件は明らかではないかもしれないし，方程式に加える逆数が増えれば増えるほど，その要件はさらに低くなるだろう．

### INCLUSION OF IRRELEVANT VARIABLES

回帰モデルが（4.12）式の通りに正しく与えられ，（4.8）式が正しいかのように推定する（つまり，いくつかの追加の変数を含める）場合，先に検討したのと同じ種類の問題が発生するように思えるかもしれないが，実際には，このようなケースは起こらない．

|  |  |
| --- | --- |
|  | (4.12) |

関連する変数のセットを省略することは，(4-8) に誤った制限を課すことと同じだと考えることができる．特に，を省略することは，いう制約のもとで(4-8)を正しく推定していないことと同じである．しかし，この誤差が単に正しい情報を使わなかっただけだとする．制限を誤ってかけると，偏った推定値が得られる．この誤差を別の見方をすると，それは誤った情報を推定に取り入れたことになる．しかし，仮に誤差は単に正しい情報を使用しなかっただけだとする．

無関係な変数を回帰に含めることは，(4-8)の推定にを課さないことと同じである．しかし，(4-8)は間違っているわけではなく，単に を組み込むことに失敗しただけである．したがって，(4-8) のの最小二乗推定量に制限があっても不偏であることを形式的に証明する必要はなく，すでに証明されている．これまでのすべての結果に基づいて，次のように主張することができる．

|  |  |
| --- | --- |
|  | (4.13) |

では，どこに問題があるのだろうか．一般的には，モデルを「オーバーフィット」させたいと思われるだろう．理論的には，正しい情報を使用しないことは常にコストがかかるというのがこの考え方の難点です．この例では，コストは推定値の精度を下げることになる．セクション4.7.1で示すように，（を省略した）短い回帰の共分散行列は，余分な変数がある場合に得られる推定量の共分散行列よりも決して大きくない．

一変数の比較を考える．がと非常に相関している場合，を誤って回帰に含めると，の推定値の分散が大きくなる．

### THE VARIANCE OF THE LEAST SQUARES ESTIMATOR

分析者がの値を選択する実験的な状況のように，回帰因子を非確率的なものとして扱うことができれば，を定数の行列として扱うことで，最小二乗推定量のサンプリング分散を導き出すことができる．また，が確率的であることを許容し，観測されたを条件として分析を行い，(4-5) から(4 6) を求めたときと同様に，Xの平均化を考慮することもできる．

(4-4)を再度用いると，次のようになる．

|  |  |
| --- | --- |
|  | (4.14) |

(ただし，) と書けることより，は外乱の線形関数であり，これからの定義では，これは線形推定量となる．前述の通り，(4-14)の第2項の期待値は0です．従って， の分布に関わらず，他の仮定の下では，はの線形で不偏な推定量となる．仮定A4により， となる．

したがって，最小二乗法による傾き推定量の条件付共分散行列は

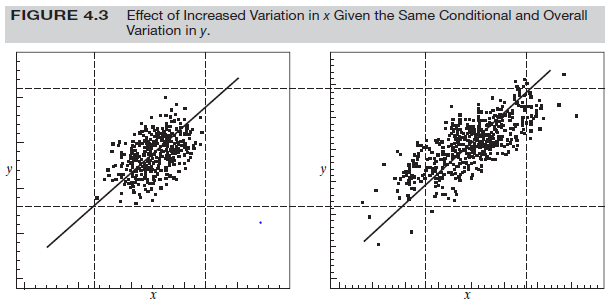
|  |  |
| --- | --- |
|  | (4.15) |

Example 4.3　 2変数回帰モデルにおけるサンプリング分散（Omitted Variable）

には定数項（1の列）と単一の回帰子のみが含まれているとする．の右下の要素は次の通り．

|  |  |
| --- | --- |
|  |  |

特にの分散の分母に注目する． の変動が大きいほど，この分散は小さくなる．例えば，図4.3の2つの回帰の傾きを推定する問題を考えてみる．より正確な結果が得られるのは，図4.3の右側のパネルのデータである．



ここで，の線形不偏推定量のクラスに関する一般的な結果が得られる．

**≪Theorem 4.2（Gauss – Markov Themorem）≫**

線形回帰モデルにおいて，回帰行列 を用いた場合，最小二乗推定量 はの最小分散線形不偏推定量である．

また，任意の定数ベクトル に対して，線形回帰モデルにおけるの最小分散線形不偏推定量はである．ただし，は最小二乗推定量である．

なお，この定理では前提条件A6の「外乱分布の正規性」は利用していない．必要なのはA1からA4だけである．この重要な定理を証明するための直接的なアプローチは， となるような線形で不偏な推定量を定義し，そのクラスの中で最小の分散を持つメンバーを見つけることである．

その代わりに間接法を使います．すでにが線形不偏推定量であることを証明した．次に，の他の線形不偏推定量を検討し，そのような他の推定量は分散が大きいことを示す．

Proof )

をの別の線形不偏推定量とし，は()行列であるとする．が不偏であれば，

|  |  |
| --- | --- |
| ただし， |  |

上式を満たす候補はたくさんある．例えば，の最初の行（あるいは，任意の行）の線形独立な行だけを使うことを考えてみる．

このとき，

ただし，はの行から形成される行列の逆行列である．

このとき，の共分散行列は

と置き換え（4.14）式を用いることで考えることが出来る．結果としては，

とする．つまり，

このとき，

ここで，

であるから，

であるはずである．したがって，

の二次形式は，

であるから，条件付き共分散行列は，の条件付き共分散行列に非負定値行列を加えたものに等しい．したがって，のすべての二次形式はの二次形式よりも大きく，これにより第一の結果が成立する．

の分散はの二次形式であり，同様に任意のについても，二次形式であることから，2番目の記述の証明は前の導出から導かれ，個々の傾き推定量がの最良の線形不偏推定量であることが証明される．(を番目の位置の1を除くすべてのゼロとする)．しかし，この定理はこれよりもはるかに広いものである．なぜなら，この結果はの要素の他のすべての線形結合にも適用されるからである．

# 第2回

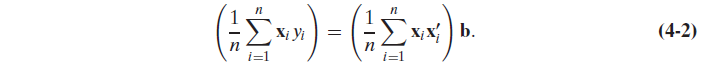
### Maximum Likelihood Estimation

これまでの話について2つの方法で最小二乗推定量の動機付けを行った．

1つ目は，最小二乗推定量が最小分散推定量と一致しているというTheorem 4.1を得た．これは，との同時分布における最小二乗推定量は，との同時分布におけるyの最小平均二乗誤差予測値の係数を模倣するという定理であった．

≪Theorem 4.1≫Minimum Mean Squared Error Predictor (P94)

を生成するデータ生成機構が，(4-1)の行列の(4-2)の推定量に大数の法則が適用されるようなものであれば，の最小期待二乗誤差線形予測値は最小二乗回帰線によって推定される．



Theorem 4.2（ガウス・マルコフの定理）では，最小二乗推定量は，最小分散線形不偏推定量であることを示していた．つまり，線形回帰モデルの仮定の下でβの最小分散線形不偏推定量であることを示している．

**≪Theorem 4.2（Gauss – Markov Themorem）≫**

線形回帰モデルにおいて，回帰行列 を用いた場合，最小二乗推定量 はの最小分散線形不偏推定量である．

また，任意の定数ベクトル に対して，線形回帰モデルにおけるの最小分散線形不偏推定量はである．ただし，は最小二乗推定量である．

どちらの結果も，前提条件A6，つまりεの分布の正規性に依存していない．この時点で感じる自然な疑問は，この仮定の役割は何かということである．

仮定の役割は2つある．

1. 正規性の仮定は，4.5節と4.6節で信頼区間の適切な終点を決定するための基礎となる．しかし，4.4.2節で中心極限定理に基づいて，外乱が正規分布に従わなくても，の漸近的な正規分布に基づいて推論することができることが分かった（Theorem 4.4）．これにより，正規性の仮定はもはや必要ないように思われますが，2つ目の結果がなければほぼこの考えは正しい．

撹乱項が正規分布している場合，最小二乗法による推定量は**最尤推定量（MLE）**でもある．最尤推定については第14章で詳しく検討しているので，この時点では簡単にしか説明しない．

最終的には，（最尤推定量）MLEであることにより，最小二乗法は一貫性があり，漸近的に正規分布する推定量の中で効率的であるということである．

これは，Gauss-Markovの定理（正式にはCram´er-Rao boundと呼ばれる）の大規模サンプル対応である．この2つの定理に共通しているのは、想定される推定量のクラスにおいて、最小二乗法による推定量が最も効率的な推定量であることを示していることである．

この2つの定理の違いは、想定する推定量のクラスにあります。

|  |  |
| --- | --- |
| **Gauss-Markov** | **ML** |
| 線形で不偏な推定量 | 正規分布した外乱に基づいた、整合的で漸近的に正規分布した推定量 |

これらは "入れ子 "になっているわけではない．

例えばMLEの結果は，不偏性（unbiasedness）や線形性（linearity）を必要としないことに注意すること．

Gauss-Markov は，正規性（normality）や一貫性（consistency）を必要とない．

|  |  |
| --- | --- |
| **Gauss-Markov** | **MLE** |
| 不偏性，線形性 | 正規性，一貫性 |

Gauss-Markovの定理は有限標本の結果であり，Cram´er-Rao bound は漸近的（大標本）な性質である．

今回の開発（紹介）で重要な点は，効率性に関するものです．効率とは，統計的推論のためにサンプルデータをどのように使用するのが最適かという問題に関連している．一般的に，ある推定量が効率的であることを証明することは，候補者を特定することなしには困難である．

Gauss-Markovの定理は，線形回帰モデルの強力な結果である．しかし，この定理は他のモデリングの文脈では対応するものがないため，線形モデルから離れると，推定量を比較するための異なるツールが必要になる．

最尤推定の原理により，分析者は推定量の漸近的効率性を主張することができるが，それは仮定した特定の分布に対してのみである．

例4.6では，正規分布の外乱を持つ回帰モデルにおいて，がMLEであることを証明している．例4.7では，回帰の外乱が正規分布に従わないとき，その結果，がMLEよりも効率が悪い場合を考える．

≪Example 4.6≫ MLE with Normally Distributed Disturbances

正規分布の外乱がある場合， は平均，分散 なので，の密度は



個の独立した観測値のサンプルに対する対数尤度は，観測された確率変数の同時密度の対数に等しい．ランダムなサンプルの場合（独立同分布と仮定する場合），同時密度関数は積になるので，データが与えられたときの対数尤度はと書かれ、密度の対数の合計になる．これは（少し操作して）次のようになる．



この関数を最大化するとの値が，との最尤推定量である．第14章でさらに検討しますが，と に関してこの関数を最大化するデータの関数は，最小二乗係数ベクトルと平均二乗残差である．繰り返しになるが，以下の結果の導出は第14章に譲る．



これは4.3.6節で出てきた内容と同じである（P101）．これにより，最小二乗推定量は最尤推定量であることがわかる．これは一貫性があり，漸近的に（正確に）正規分布し，正規性の仮定の下では，定理14.4（調べよう）により，漸近的に効率的である．

MLEの特性は，観測された確率変数に仮定された特定の分布に依存することに注意することが重要である．もし，に非正規分布が仮定され，がMLEでないことが判明した場合，最小二乗法は効率的ではないかもしれない．次の例で説明する．

≪Example 4.7≫ The Gamma Regression Model

Greene (1980a)は，非対称に分布する外乱を含む回帰モデルの推定について考察している．



ここで，はB.4.5節のガンマ分布［(B-39)参照］であり，は外乱の標準偏差である．このモデルでは（定数項を含まない）傾き係数の最小二乗推定量の共分散行列は



一方，最尤推定量（これは最小二乗推定量ではない）については



しかし，非対称パラメータについて，この結果は最小二乗法による推定量と同じになる．つまり非対称な外乱分布を考慮した推定量の方が漸近的には効率的であると結論づけている．

例4.7のモデルと多少似ているもう一つの例は，第18章で開発されたStochastic Frontier Model（すときゃスティック・フロンティア・モデル）である．特に，この2つのケースでは，外乱の分布が非対称である．最尤推定量はこの点を特に考慮して計算されているが，最小二乗推定量は，回帰線の上と下の観測値を対称に扱う．

この違いが，この2つのモデルに対するMLEの漸近的な優位性の源となっている．

## Interval Estimation

区間推定の目的は，パラメータの最良の推定値を，その推定値に付随する不確実性を明示して示すことである．パラメータθの推定のための一般的なアプローチは，次のようになる．



(ここでは，対象となる区間がを中心に対称的であると仮定している．) サンプリングのばらつきの範囲が（非）確実性の度合いを示すべきであるという論理に従い，論理的な両極端を検討する．

推定しているパラメータの真の値がの範囲にあることを絶対（100％）に確信することができる．もちろん，これは特に有益な情報ではない．もう一方の極端な例としては， の範囲に何の確信も持たない（0％）ことが挙げられる．真のパラメータ値に正確に当たる確率はゼロと考えるべきである．

ポイントは，(4-13) の区間にパーセントという必要な信頼性（確率）を付けられるように，αの値を（0.05または0.01が一般的ですが）選択することである．この目的のために，外乱が正規分布しているという仮定A6から離れる．そして，その仮定を緩和し代わりに推定量の漸近的な正規性に頼ることにする．

### FORMING A CONFIDENCE INTERVAL FOR A COEFFICIENT

(4-18)より、 となる．これにより，bの任意の特定の要素，例えばに対して



ここでは(XX) の 番目の対角要素を表す．この変数を標準化（標準正規分布に従うようにする）すると，次のようになる．



は， ，，，の関数だが，モデルのパラメータやデータのいずれにも関与しない分布を持っていることに注意するように．従来の95％信頼水準を用いると，となる．簡単な操作で，次のことが分かる．



これは、ランダム区間 にが含まれる確率を述べているのであって，が指定された区間に含まれる確率ではないことに注意してください．もし，95%ではなく他の信頼度を使いたい場合は，(4-39)の1.96を適切なに置き換える．

という表記は，標準正規変数に対して、となるの値を表している．したがって，で，これはα＝0.05に相当する）

(4-39)のような信頼区間が得られるが，がわからないため，信頼区間が運用できないという複雑な問題がある．回帰からを使うのが自然に思える．これは確かに適切なアプローチである．



(4.40)は，標準正規ベクトルの冪乗二次形式である．したがって，自由度が に等しいカイ二乗分布となる（この結果の証明はB11.4節参照）．(4-40) のカイ二乗変数は(14)の標準正規変数とは独立である．これを証明するには，次のことを示せばよい．



は、と独立である．B.11.7項(定理B.12)では，標準正規ベクトルにおける一次形式と冪乗二次形式 の独立性の十分条件は，であることが分かった．をとすると，ここでの条件は = 0であることがわかる．この一般的な結果は，回帰分析における多くの検定統計量の導出の中心となる．

≪Theorem 4.6≫ Independent of and

が正規分布に従えば，最小二乗係数推定量は，残差ベクトルeから統計的に独立しており，したがって を含むeのすべての関数から独立している．



したがって、この比は個の自由度を持つ分布となる．(4-41)の結果は、の代わりに を使い，分布も標準正規分布ではなく，自由度がの分布という重要な分布を使っている点で，(14)とは異なる．これにより，の信頼区間は次のようにして形成される．



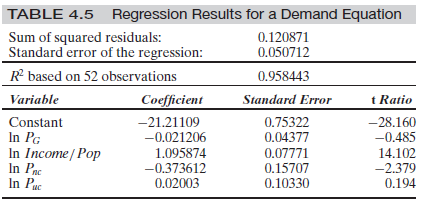
ここではt分布からの適切な臨界値である．ここで，極めて重要な統計量の分布は，までのサンプルサイズに依存するが，パラメータやデータには依存しないことが分かる．(4-42)の実用的な利点は，未知のパラメータを含まないことである．の信頼区間は(4-42)に基づいて求めることができる．

≪Example 4.8≫ Confidence Interval for the Income Elasticity of Demand for Gasoline

例4.2および例4.4で説明したガソリン市場のデータを用いて、52の観測データから次のような需要方程式を推定しました。



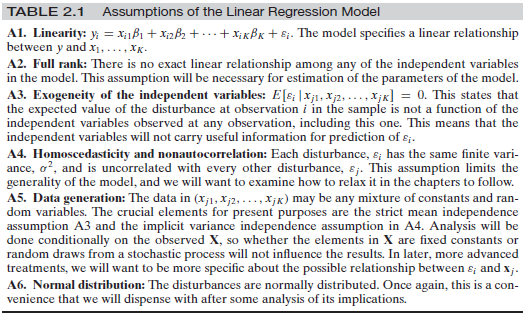
モデルパラメータの最小二乗法による推定値、標準誤差、t比を表4.5に示します。



所得弾力性の信頼区間を形成するためには， の自由度を持つ 分布からの臨界値が必要です．95パーセントの臨界値は2.012です。したがって、の95%信頼区間は,

# Data Problems

ここまでの分析では，手元のデータとがよく測定されており表2.1のモデルの仮定と，基礎理論で説明される変数に対応していることを前提としてきました．ここでは「現実の」観察された非実験的データが仮定を満たさない場合のいくつかの方法を検討します．仮定が満たされない場合，一般的にモデルのパラメータの推定値の性能に影響を与え，その影響は良いものではありません．ここで検討するケースは以下の通り．



* **多重共線性（Multicollinearity）**

フルランクの仮定A2は満たされていますが，ほとんど失敗しています．(ほとんど」というのは程度の問題であり、時には解釈の問題でもあります） 多重共線性は，推定量の不正確さにつながりますが，推定に系統的な偏りがあるわけではありません．

* **欠損値（Missing values）**

やのギャップは無害な場合もある．多くの場合，分析者はそれらを単に無視して，サンプルの完全なデータを使用することができます（そして，そうすべきです）．他のケースでは，調査対象の結果に関連する理由でデータが欠損している場合，問題を無視すると推定量が不整合になる可能性があります．

* **測定誤差（Measurement error）**

所得や教育に関する個人のデータなど，データはモデルに登場する理論的な構成要素に不完全にしか対応していないことがよくあります．測定誤差は決して良いものではありません．最も害の少ないケースは，説明変数の測定誤差です．この場合，少なくとも合理的な仮定の下では，データが正確に測定された（残念ながら仮定の）場合に比べて，データに対するモデルの適合度が低下することになります．被説明変数の測定誤差は悪質であり，修正が困難な推定における恒常的なバイアスを生み出します．

## 多重共線性（Multicollinearity）

ガウス・マルコフの定理は，すべての線形不偏推定量の中で，最小二乗推定量の分散が最も小さいことを示しています．この結果は有用ですが，絶対的な意味で最小二乗推定量の分散が小さいことを保証するものではありません．

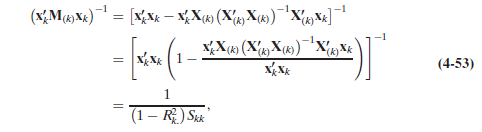
例えば，2つの説明変数と定数を含むモデルを考えてみましょう．どちらかの傾きの係数に対して



2つの変数が完全に相関している場合は，分散は**無限大**になります．目標変数（regressors）の間に正確な線形関係がある場合は，データではなくモデルの仮定に重大な障害があると考えられます．より一般的なケースは，変数が完全ではないものの高度に相関している場合です．この場合，回帰モデルはその仮定された特性をすべて維持しますが，潜在的に深刻な統計的問題が発生します．応用研究者が直面する問題は，回帰変数が完全ではないが高度に相関している場合で，次のような症状がある．

* データのわずかな変化によって，パラメータの推定値が大きく変動します。
* 係数の標準誤差が非常に大きく，有意水準が低い場合があります。係数が共同で有意であり，回帰の決定係数がかなり高くても，標準誤差が非常に大きく，有意水準が低い場合があります．
* 係数の符号が間違っていたり，大きさが不自然だったりします。

便宜上，データ行列は，定数と，平均値からの偏差で測定される個の他の変数を含むと定義する．を番目の変数とし，を他のすべての変数（定数項を含む）とする．すると，，逆行列の番目の対角要素は

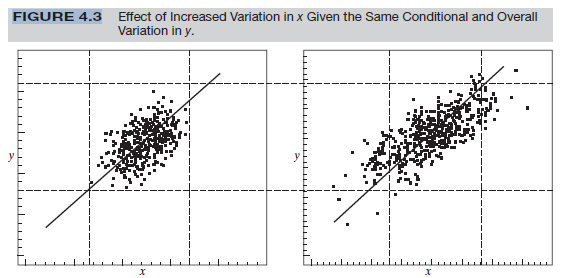


ここで、.はを他のすべての変数に回帰させたときのである．重回帰モデルでは，番目の最小二乗係数推定量の分散は，この比率の倍となる．つまり，モデル内の他の変数との相関性が高いほど（集合的に），その分散は大きくなるということです．最も極端なケースでは，が他の変数の線形結合として書かれ，となるため，分散は無限大になります．その結果，



は，番目の最小二乗係数推定量の精度の3つの要素を示しています．

* 他の条件が同じであれば，と他の変数の相関が大きいほど**多重共線性**のため，分散が大きくなります．
* 他の条件が同じであれば、の変動が大きいほど，分散は小さくなります．この結果は，図4.3に示されています．
* 他の条件が同じであれば，回帰の全体的な適合性が高ければ高いほど，分散は低くなります．この結果は，の値が低いことから導かれます．この意味をまだ理解していませんが，図4.3の右図と同じ図を想像して，すべての点を回帰線の近くに移動させることで，この意味を示唆することができます．

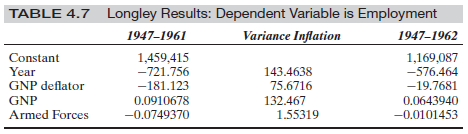


非実験データが直交することはない（相関が0になることはない）ので（），ある程度の**多重共線性**は常に存在します．では，どのような場合に**多重共線性**が問題となるのでしょうか．つまり，この相互相関によって推定値の分散が悪影響を受け，「懸念」する必要があるのはどのような場合なのでしょうか？いくつかのコンピュータパッケージは，診断統計量として回帰の各係数の分散インフレーション係数（**variance inflation factor** ：VIF），を報告します．見てわかるように，ある変数のVIFは，その変数がモデル中の他の変数に直交していないという事実に起因するの増加を示します．

に特化したもう一つの指標は，の条件数（Condition Number）であり，これはの最大の特性根（単位長さになるように各列をスケーリングした後のもの）と最小の特性根の比の平方根である．[Belsley, Kuh, and Welsch (1980)]では，20を超える値は問題があることを示唆している．(この 例4.11のLongleyデータの条件数は15,000を超えている！)

≪例4.11 Longleyデータの多重共線性≫

付録表F4.2のデータは，J. Longley (1967)がコンピュータプログラムによる最小二乗計算の精度を評価する目的で作成したものである（このデータは現在も広く使用されている）．(Longleyのデータは**多重共線性**が強いことで有名である．例えば，データセットの最後の年を見てください．最後の観測値は異常ではないように見えます．しかし，表4.7の結果は，定数と他の変数に対する雇用の回帰から，この1つの観測値を削除した場合の劇的な効果を示している．最後の係数は600%上昇し，3番目の係数は800%上昇しています．



多重共線性を発見し，それに対処するための戦略がいくつか提案されています．多重共線性の「問題」は，情報が不足しているために発生するという見解のもと，1つの提案は，より多くのデータを得ることです．アナリストが最初からそのような追加情報を持っていたならば，このような状況になる前にそれを使うべきだったと主張する人もいるかもしれません．しかし，情報が多ければ観測回数が多いというわけではありません．

明らかに実用的な解決策（そして間違いなく最も頻繁に使用されている）は，問題の原因と疑われる変数を回帰から削除することです．つまり，「問題のある」変数はモデルに現れないという，おそらく誤った仮定を回帰に課すことです．この方法では，セクション4.7.2で議論する仕様の問題に遭遇します．

削除された変数が実際にモデルに含まれている場合（その係数がゼロではないという意味で），残りの係数の推定値には偏りがあり，場合によっては重大な偏りが生じることになります．一方，オーバーフィッティング（大きすぎるモデルを推定しようとすること）は一般的なエラーであり，過剰に指定されたモデルから変数を削除することは有益な場合があります．

診断ツールを使って多重共線性を「検出」することは，悪いモデルと悪いデータを区別しようとしていると考えられます．しかし，実際の問題は，データが対立しているように見える事前の意見に起因するだけです．多重共線性が推定値に悪影響を与えているという所見は，この影響がなければ，すべての係数が統計的に有意で正しい符号を持っていることを示唆しているように思えます．

もちろん，このような状況である必要はありません．ある変数がモデルの中で重要でないことをデータが示唆している場合，理論はともかく，研究者は最終的にその理論にどれだけ強くコミットするかを決めなければなりません．多重共線性の「救済策」として提案されるものは，理論をデータに押し付けようとするものかもしれません。

## プレテスト推定 (PRETEST ESTIMATION)

「多重共線性」と思われる問題への対応として，問題を起こしていると思われる変数を回帰から削除したいという誘惑には，しばしば抵抗があります．この「戦略」は，分析者に微妙なジレンマをもたらします．分割された重回帰を考えてみましょう．



をのみに回帰すると，推定量に偏りが生じます．



この推定値の共分散行列は



(この分散はの分散であり，の分散ではないことに注意してください) が実際にはゼロではない場合，の()への重回帰において，の平均の分散は



ただし，



または，



2つの共分散行列を比較します．逆数を比較する方が簡単です．[参照 結果(A-120)参照)] したがって



は，非負定値行列です．このことから，の分散はの分散よりも大きくないことがわかります（その逆数は少なくとも同じ大きさなので）．このことから，は偏っているが，その分散は偏っていない推定量の分散よりも大きくなることはない．現実的なケース（つまり，がゼロでない場合）では，実際にはもっと小さくなります．平均値からの偏差として測定される2つの変数を用いた単純な回帰から，有用な比較を得ることができます．

前項の結果は，応用研究者にちょっとしたジレンマを与えています．これは，モデルの仕様を探す際に頻繁に発生する状況です．研究者がモデルに入れるべきだと疑っている変数があり，それが多重共線性の問題を引き起こしている場合，分析者は関連する変数を省略するか，それを含めてその（そして他のすべての変数の）係数を不正確に推定するかの選択に直面します．この場合，とという2つの推定量の選択となります．実際には，研究者が通常行うことは，第3の推定量を生み出すことになります．問題のある変数を暫定的に含めるのが一般的です．

その比が十分に大きければ，それは保持され，そうでなければ破棄される．この第3の推定量は，プレテスト推定量と呼ばれています．プレテスト推定量について知られていることは，心強いものではありません．確かに偏りがあります．どの程度の偏りかは，未知のパラメータに依存します．分析結果によると，研究者が使用する可能性が最も高い場合，プレテスト推定量は3つのうち最も精度が低いことが示唆されています．導き出される結論は，一般的なルールとして，方法論は多重共線性のための場当たり的な救済策を含む推定戦略から遠ざかるということです.

## Principal Component

多重共線性を"減らす "ために提案されている方法[例えば、Gurmu, Rilstone, and Stern (1999) を参照]は，元の変数の線形結合として構築された少数，例えばL個の主成分を使用することである [Johnson and Wichern (2005, Chapter 8)を参照]（その仕組みは例題4.12に示されている）．このアプローチを使用することに対する議論は，という元の仕様が正しい場合，をの列の線形結合のいくつかの小さなセットに回帰したときに，何を推定しているのかが不明であるということである．

の主成分の集合に対して，y を に回帰して を求めると、 となる．(証明は演習で考える) 経済学の文脈では，に解釈があるならば、にも解釈があるとは考えにくい．(例えば，弾力性から所得弾力性の2倍を引いた 所得弾力性の2倍を引いたものをどう解釈するかなど)

このオーソドックスな解釈は，多重共線性に対処するための機械的な装置が，解釈できないような係数の混合物を生み出すことについて，分析者に注意を促します．しかし，モデルがいくつかの測定された変数の混合物を含む可能性のあるプラットフォーム上に構築されている状況もあります．例えば，曖昧に定義された「能力」を含む回帰モデルに興味があるかもしれません．測定された対応物として，アナリストは一連のテストの標準化されたスコアを持っているかもしれませんが，そのどれもがモデルの文脈では特定の意味を持ちません．この場合，測定されたテストスコアの混合物が，基礎となる変数の好ましい代理となるかもしれません．例4.12の研究は，もう一つの自然な例です。

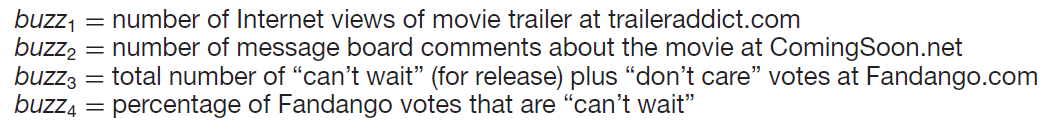
≪例4.12 映画の成功予測≫

映画の興行成績を予測することは，計量経済学者が好んで行う課題である．[例えば、Litman (1983), Ravid (1999), De Vany (2003), De Vany and Walls (1999, 2002, 2003), Simonoff and Sparrow (2000)などを参照のこと）] 伝統的な予測式は次のような形になる．



決定係数は0.4のオーダーが一般的です．このようなモデルが比較的強力であるにもかかわらず，ハリウッドでは「誰にもわからない」というのが常識です．映画の成功には非常に大きなランダム性があり，それを確実に予測できると考える人はほとんどいないのです．

Versaci（2009）は、「インターネット・バズ」という新しい要素をモデルに加えました．インターネット上の話題とは，「RottenTomatoes.com」，「ImDB.com」，「Fandango.com」，「traileraddict.com」などのよく知られたウェブサイトにおけるインターネット上のトラフィックと関心を指すと漠然と定義されています．これらのサイトは、それ自体ではインターネットの話題性を定義するものではありません．しかし，これらのウェブサイトでの活動を総合するとこれらのサイトでの活動は，例えば映画公開の3週間前になると，今後の成功を予測するのに役立つかもしれません。．Versaciのデータセット（表F4.3）には，2009年に公開された62本の映画のデータが含まれており，4つのインターネット・バズ変数（Internet buzz variables）が含まれています。

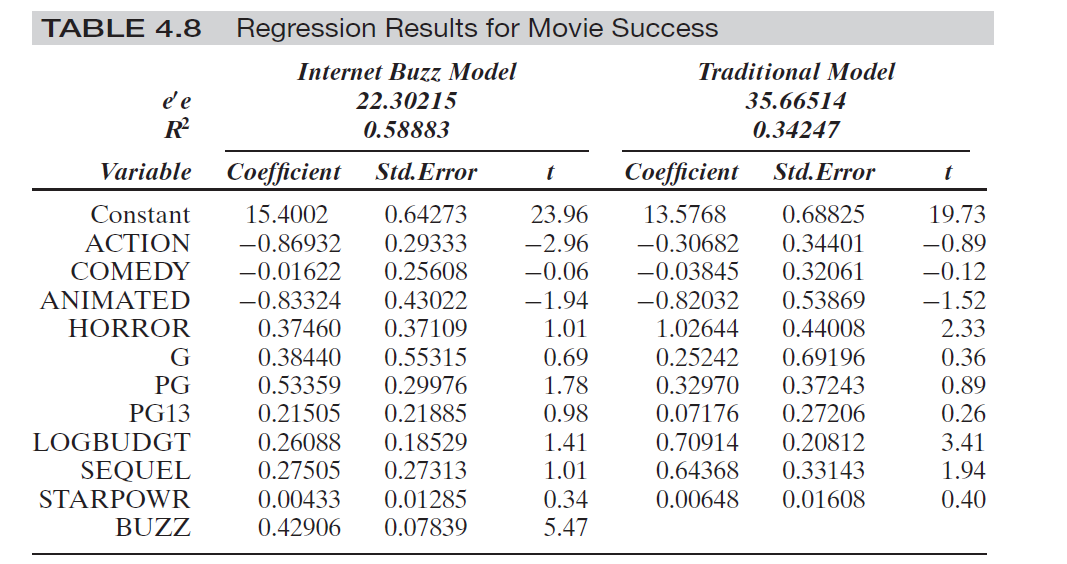


以下のように，これらを1つの主成分に集約しました．まず，スケール効果を除去するために，buzz1 - buzz3のログを計算しました．次に，4つの変数を標準化して，は元の変数からその平均値を引いて，その標準偏差で割ったものを含みます．結果として得られた62×4の行列（）をとします.

が標本相関行列です．最大の特性根に関連するVの特性ベクトルをとします．第1主成分（4つの変数の変動のほとんどを説明するもの）はです．(根は2.4142, 0.7742, 0.4522, 0.3585なので，第一主成分は2.4142/4，つまり変動の60.3%を説明します．

表4.8は，2009年に公開された62本の映画を対象とした回帰分析の結果です．インターネットでの話題性が，回帰の予測力を大幅に高めていることがわかります．回帰のは0.34から0.58と2倍近くになっていますが,

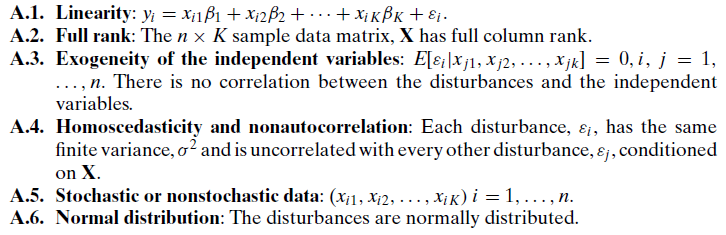
これはインターネット上の話題がモデルに加わったことによるものです. 第5章で説明するように，話題性は成功の高い「有意な」予測因子でもあります．



# 第4回

## Assumptions of Extended Model

第2章と第4章で説明した線形回帰モデルの前提条件は以下の通り．



ここでは、

という重要な結果を維持します。しかし、回帰モデルの基本的な前提条件が変わりました。

まず、A.3（との間には相関がない）は、新しい仮定の下では



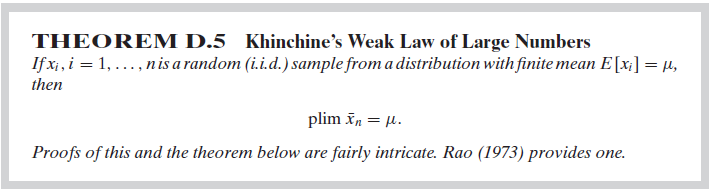
前提条件A.I3は、リグレッサが外乱の期待値に関する情報を提供するようになったことを意味すると解釈する。A.I3の重要な意味は、外乱とリグレッサが相関するようになったことである。

前提条件A.I3は、 に対して



もし、データが ”well behaved” であれば、定理D.5 (Khinchineの定理(P1110)）を適用して、次のように主張することができます。

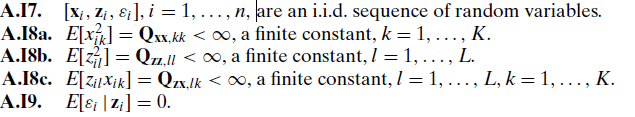




今までのモデルでは，を仮定していた。 (8-3)の意味するところは、回帰因子はもはや外生的ではないということである。ここで、2つの特性を持つ変数Zが追加されたと仮定します。

1. **外生性**…これは，撹乱項と無相関である．
2. **妥当性**…これらは独立変数であるXと相関関係を持つ．

これらの概念は、今後、公式化していく予定です。我々のモデルでは、これら2つの特性を持つ変数を道具変数（Instrument Variable）としています。



≪復習（Instrument Variable）≫

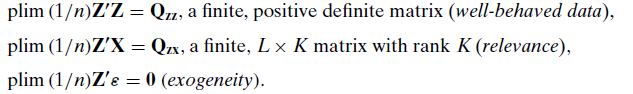
instrumental variablesの手法は、以下のような推定戦略に基づいて開発されている。

(8-1)のモデル



において，個の変数がと相関している可能性があると仮定する。同様に、がと相関しているがとは相関していないような個の変数のセットが存在するとする。おなじみの最小二乗法による推定では、を一貫して推定できません。しかし、 との間に相関がないと仮定すると、 の間の仮定された関係を使用して、の一貫した推定値を構築できる可能性がある一連の関係を意味します。

後の時系列モデルの研究では、仮定A.I7を緩和することが重要になります。有限の平均値は、A.I8b. から導かれる。4.4節と同様の分析を行うと、次のようになります。



これまでの回帰モデルの記述では、，の特殊なケースを仮定してきた。仮定A.I7を無視する必要はなく、今後も真であり続けるかもしれませんが、この特殊なケースでは無相関（Irrelevant）になります。

≪前に出た例の話≫

ここでは、とする。方程式の右辺の変数と同じ数のInstrument Variableがあると仮定します。序論と例題8.1で、(8-1)の右辺でを個の外生変数のセットと個の内生変数のセットに分割したことを思い出してください。例8.1の内生的治療効果に関するKreuger and Dale (1999)の研究では、式中に単一の内生変数である治療ダミー変数があります。このような場合、個の変数はの道具変数の一つとなり、個の残りの変数はと同じではない他の外生変数となるということです。

通常の解釈では、これらの変数は「の道具」であり、変数はそれ自体の道具であると考えられる。

例を続けると、内生的治療効果モデルの行列には、の列と、治療ダミー変数のための追加の道具変数が含まれます。需要と供給の連立方程式モデルでは、

内生的な右辺の変数は　=価格

外生的な変数は　()

このモデルでは、操作変数のセットは

ではないかと（正しく）考えるかもしれません。変数間の基本的な関係については、この直感的な理解が信頼できるガイドとなるだろう。しかし、理由はすぐに明らかになりますが、統計的にはを全体の道具として扱う必要があります。

Instrument Variableの使用については、2つ目の微妙なポイントがありますが、それは以下で明らかになります。「関連性の条件」は、実際には条件付き相関の記述でなければならない。もう一度、治療効果の例を考え、が治療ダミー変数のための問題の道具変数であるとします。形式的には、との条件付き相関がゼロでないことが要求されます。これを見る1つの方法は、射影の観点からです。

のへの回帰におけるの係数がゼロでない場合、道具変数は関連性があります。直感的には、はモデルにすでに含まれている変数では提供されないの動きに関する情報を提供しなければならない。

## Estimation

セクション8.2の一般的なモデルでは、最小二乗法で得られた有用な結果のほとんどが失われます。最小二乗法の意味を考え、この拡張モデルでのβの代替推定量を構築します。

### Least Square

最小二乗法による推定値は、もはや不偏ではありません。

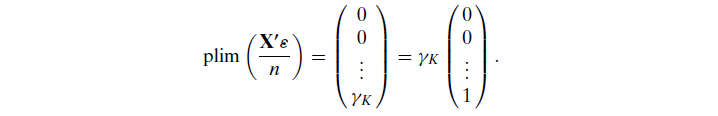


つまり、Gauss-Markovの定理が成立しない。また、この推定も成り立たない。



(最小二乗法の不整合は、内生変数の係数に限ったことではない。)

これを確認するために、(8-4)を前述の治療効果の例に当てはめてみましょう。この場合、の最後の変数以外はと無相関です。



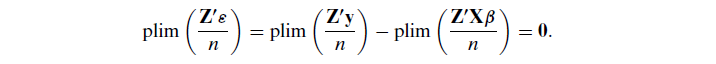
この特別なケースでは、（8-4）の結果は、次のようになります。



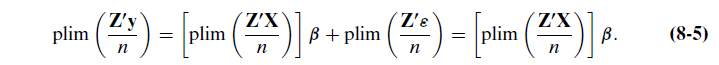
の最後の列の要素のどれかがゼロになると予測する理由はありません。つまり、の変数のうち1つだけがと相関しているにもかかわらず、内生変数の係数の推定値だけでなく、のすべての要素が矛盾しているということです。この効果はスミアリングと呼ばれ、1つの変数の内因性による不整合が、すべての最小二乗法の推定量に渡ってスミアリングされます。

### THE INSTRUMENTAL VARIABLES ESTIMATOR

であり、すべての項が有限の分散を持つため、次のようになります。

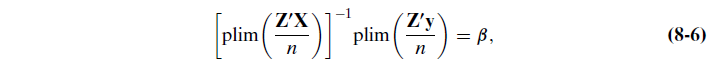


従って、



はと同じ数の変数を持っていると仮定しました。例えば、消費関数において、 のとき、

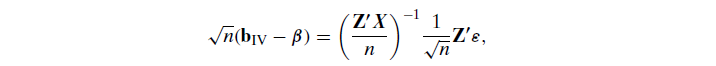
であると仮定します。のランクはであると仮定したので、は正方行列である。このとき、以下が成り立つ。



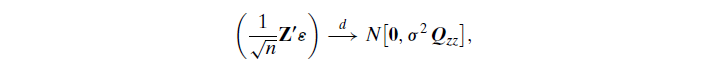
となり、道具変数推定法が導かれます。



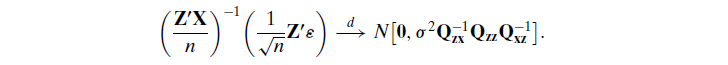
我々はすでにが一貫していることを証明した。次に、漸近的な分布について、4.4.2節と同じ方法で説明する。まず、



は、と同じ限界分布を持っています。の分析は、4.4.3項のになるので、次のようになります。



また、



このステップで、次の定理の導出が完了します。

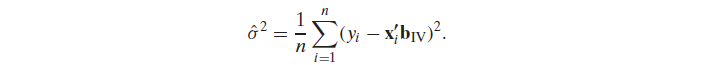
≪定理8.1≫

仮定A.1、A.2、A.I3、A.4、A.5、A.I7、A.I8a-c、A.I9がすべてについて成立し、が instrumental variablesの有効なセットである場合、 instrumental variables estimatorの漸近分布 は



*ただし、and*

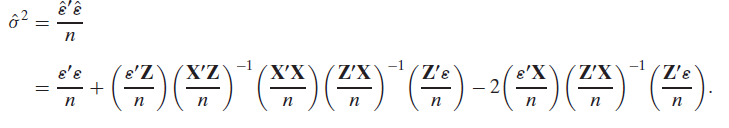
漸近的共分散行列を推定するためには、σ2の推定量が必要です。自然な推定値は



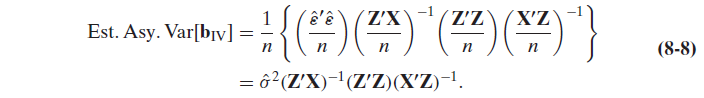
ここでの結果はすべて漸近的なものであり、どのような場合でもは不偏ではないので、自由度の補正は不要です（それでも、ほとんどのソフトウェアでは自由度の補正を行うのが標準的です）。残差のベクトルを次のように書きます。



を代入し、整理すると、が得られる。これで



以前、確率極限（Probability Limit）の積の結果を応用して、このような式の確率極限を求めることができることを発見しました（少しの操作で）。導出を繰り返す必要はありませんが、第1項のおかげで、はの一貫した推定量であることがわかります。また、第2項と第3項はゼロに収束します。導出を完成させるために、我々はを推定する。を



# 8.4.2. A TEST FOR OVERIDENTIFICATION

IV推定量を選択した動機は，効率性ではない．推定量は一貫性があるように構成されており，効率は考慮されていない．第13章では，効率的なモーメント法推定の問題を再検討する．

2SLSがすべての個の測定器を最も効率的に使用しているという観察結果は，の列の個の線形結合を使用する推定量のクラスにおける推定量の効率性のみを立証する．

IV推定量は直交条件に基づく．



これに対応するサンプルは以下のモーメント方程式である．



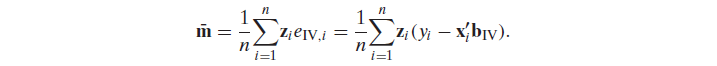
の場合の解は，これまで見てきたように である．の場合は，単一の解はなく戦略として2SLSにたどり着いた．推定は依然として(8-13)に基づいている．しかし，サンプルの対応は個の未知数に個の方程式となり、(8-13)は解がない．

それにもかかわらず，モデルの仮説の下では，(8-12)は真のままである．

私たちは，追加の制約条件をサンプルの証拠によって支持されるかもしれないし，支持されないかもしれない仮説と考えることができる．

モーメント方程式の過剰は，モデルの過同一化を検定する方法を提供する．検定は(8-13)に基づいて行われる．(8-13)はで評価されると，のときにゼロにはならないが，(8-12)の仮説はまだ真であるかもしれない．

検定統計量は**Wald統計量**になる．(5.4節参照) 標本統計量は (8-13)とIV推定量に基づく標本統計量は，



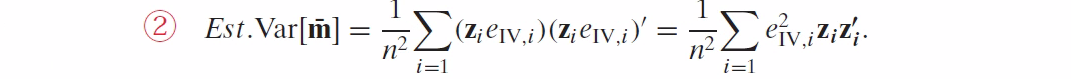
Wald統計量は，



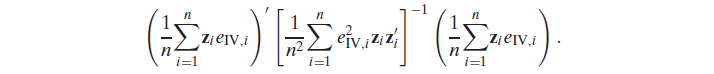
この構造を完成させるには，分散の推定量が必要である．その方法は2つある．モデルの仮定の下で，以下のものはの標本推定量を用いて容易に推定することができる．



また，(8-12)に基づいて推定することもでき，その場合，適切な推定値は次のようになる．



これらの2つの推定量は，有限のサンプルでは数値的に異なるが，これまでの仮定のもとでは，（を掛けた）両方とも同じ行列に収束するので，選択は重要ではない．現在の実務では，2番目の推定量が好まれている．Wald統計は，次のようになる．



残りの詳細は，自由度の数である．この検定は，個のモーメント方程式の失敗のみを検出できるので，それは2次形式の順位である．統計量の限界分布は，個の自由度を持つカイ2乗である．

Example 8.8. 労働供給方程式の過同一化

例8.5では、働いた週数に関する方程式のパラメータの2SLS推定値を計算した。この推定量は以下の2つである．





過剰識別制限が1つある．表8.1の2SLSの結果に基づくサンプルモーメントは，表8.1の2SLSの結果に基づく標本モーメントは，



自由度1の場合，カイ二乗統計量は1.09399となる．また，最初に提案した分散推定量を用いると，統計量は1.05241となる．どちらも95%の臨界値である3.84を大きく下回っているので，過同一化の仮説は棄却されない．

私たちは，このテストの最終的な意味を指摘する．モデルの基礎理論に基づいて，過同一化テストはある特定の道具変数に関連しており，他の道具変数には関連していないと結論づけることが出来る．例えば，RainfallとInputPriceという2つの需要方程式の道具を用いた市場均衡の例では，Rainfallは明らかに外生的なので，過同一化の制限が棄却されるとInput-Priceが有効な道具として排除されてしまう．

しかし，この結論は適切ではない．このテストは，(8-12)の要素のうち1つ以上が0でないことを示唆しているだけである．このテストでは，示唆するものではない．

# WEAK INSTRUMENTS

これまでの分析では，IV推定の「同定」条件、すなわち、A.I9の「外生性の仮定」に焦点を当ててきた．



関連性（releavance）を前提として



を仮定すると，一貫したIV推定量が得られる．

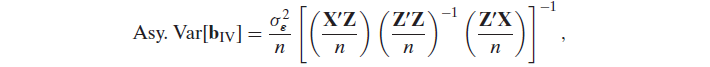
絶対的には、(8-28)があれば(8-29)は整合性を主張するのに十分である．このように，研究者はIV推定量を構築する際に解決すべき決定的な問題として外因性に注目してきた．しかし，関連性の条件にもっと注意を払う必要があると主張する文献も増えてきた．

厳密に言えば，(8-29)は我々が主張する漸近的な結果には確かに十分であるが，(8-29)がかろうじて成立するだけの「弱い測定器」の一般的なケースは，かなりの批判を集めている．

現実的には，道具（Instrument）が右辺の変数とわずかにしか相関していない場合，つまりがゼロに近い場合に「弱い」と言える．(これを理論的に定量化するのは，セクション10.6.6でこの問題を再検討するとき)．

研究者たちは，これらのケースを検討し始め，いくつかのケースでは矛盾した経験的な結果を説明することができた．

表面的には，弱計測器の問題はIV推定量の漸近的共分散行列に現れる．



上式は，楽器が弱いときには「大きく」なり，他の条件が同じであれば楽器が弱いほど大きくなります．

しかし，問題はそれだけではない．Nelson and Startz (1990a,b) と Hahn and Hausman (2003) は，2つの意味を挙げています．

1. 二段階の最小二乗推定量は，矛盾していることが知られている通常の最小二乗推定量にひどく偏っていること．
2. (ii)標準的な一次漸近法（前に使ったようなもの）では，統計的推論のための正確な枠組みを与えることができないこと．

このように，問題は単に精度が低いというだけではない．また，この問題は「小サンプル問題」をはるかに超えていることを示す証拠が少なくともいくつかある．（[Bound, Jaeger, and Baker (1995)]を参照してください。）

現在の研究では，instrumental variables の弱点を検出するためのいくつかの処方箋を提供している．単一の内生変数（と相関のある）の場合，標準的なアプローチは，2段式の第1段階の最小二乗回帰に基づいている．

回帰におけるすべての係数がゼロであるという仮説を検定するための従来の統計は，道具変数が弱いという「仮説」を検定するために使用される．



統計量が10未満の場合は問題があることを示されている．[この特定の検定の動機付けについては，Nelson and Startz (1990b)，Staiger and Stock (1997)，および Stock and Watson (2007, Chapter 12)を参照のこと]．

モデル中に複数の内生変数がある場合，変数間の共線性が結果に影響する可能性があるが，どちらのテストにも現れないので，このテストを使ってそれぞれを別々にテストすることは十分ではない．Shea (1997) は，使用可能な4段階の多変量解析手順を提案している．Godfrey (1999)は，驚くほど簡単な代替計算方法を導き出した．内生変数 について，Godfrey統計量は2つの推定量OLSと2SLSの推定分散の比率である．



ただし， は のk番目の対角要素であり，も同様に定義される．

スケーリングを行うと，統計量は次のようになる．



ここで，上付き文字は逆行列の要素を示している．

F統計は，この指標に基づいてが定数項を含むと仮定すると

となる．

弱い測定器のテストは，仕様のテストでもなければ，モデルを構築するための建設的なテストでもないことに注目してほしい．むしろ，研究者が例えば，仮定されたヌルモデル仕様の下での分布のような，馴染みのある漸近的な結果では特性がうまく表現されない信頼性の低い統計量に基づいて推論することを避けるための戦略です．

いくつかの拡張機能が興味を引く．Hahn and Hausman (2002)やKleibergen (2002)では，他の統計手順が提案されている．

また，1つ以上の内生変数がある場合にも興味がある．この問題については，セクション10.6.6で再度検討し，モデル化のフレームワークを連立方程式モデルとする．

このセクションの厳しい結果は，IV推定量に疑問を投げかけている．かなり狭い範囲ではあるが，セクション10.6.4で述べた「モーメント」フリーのLIML推定量が代替推定量となる．もう一つの方法は，おそらくやや魅力的ではないかもしれませんが，最小二乗法に戻ることである．OLS 推定量に利点がないわけでは無い．

OLS推定量の漸近的分散は，IV推定量の漸近的分散よりも明確に小さくなります。





(証明は練習問題で)これまでの結果を考えると，はるかに小さくなる可能性がある．しかし，OLSの推定値は矛盾している（8-4式を見よ）．



平均2乗誤差の比較ではAのOLS推定量とBのIV推定量のどちらがより正確かは不明である．





計量器が弱い場合の自然な対応策は，モデルから内生変数を削除するか，計量器セットを改善することである．これらはいずれも仕様上の問題である．手元にあるデータと仕様の枠組みの中での推定戦略という意味では，厳密にはOLSが好ましい戦略となる余地がある．