第 5 章　さらに進んだテクニック

この章では最小二乗法をそのまま適用するのが問題の場合を扱う。最小二乗法はある仮定のもとで統計上望ましい性質を持っている。のぞましい性質とは以下のものである。

不偏性

不偏性とは推計された係数の期待値が、母集団の真の値と等しくなることを示している。

有効性（効率性）

有効性とは、さまざまな推定値の中で、分散が最小になるように推計されたものであることを表している。最小二乗法による推定値は、ガウス＝マルコフの定理により、線形推定値の中で最小の分散を持つことが知られている。有効性と効率性は同じ意味で使われる。

一致性

データ数が多くなるにつれて、推定量が真の値に近づいていく性質を一致性と呼ぶ。一致というよりは近づいていくというイメージである。一致性とは、「小標本では望ましいかどうかはわからないが、大標本であればその推計値を使うことに意味がある」ことを表す性質だ。

最小二乗法のバリエーション

最小二乗法が望ましい推計値であるためには次の条件を満たす必要がある。その仮定とは次の 5 つであり、それが満たされない場合通常の最小二乗法は使えない。

仮定が満たされていない場合の解決法をまとめたのが右側の列である。さまざまな最小二乗法のバリエーションに対して、通常の最小事情法は OLS(Ordinary Least Squares)と略されることが多い。

線形でない式

推計しようとした方程式が、線形の形でない場合がある。しかし、たいていの式は説明変数や被説明変数を変形することで線形推定に置き換えられる。両辺の対数をとる対数線形などが代表的なものだ。

説明変数の変形などによっても線形推定に置き換えられない場合は、最小二乗法以外の推計法を使うことになる。このケースで最もよく使われるのが最尤法だ。

経済モデルとしては、定数項のない回帰式を想定することができる。この場合も残差の二乗和を最小にすることで、係数を求めることはできる。しかし、Y、X の平均が等しくない限り、「残差の平均はゼロ」という最小二乗法の仮定は満たされない。

不均一分散

　最小二乗法の仮定では、想定しているモデルの誤差が時間やサンプルを通じて一定であるとしている。次のような式を想定する。誤差項である ut の散らばり具合がサンプルを通じて一定であるという仮定である。この仮定は均一分散と呼ばれる。

　不均一分散とは、その仮定が満たされない場合で、推計した係数の分散が最小にならないことだ。

　解決法としては、まずサンプルを通じて大きさに差がないようなデータに変換することである。人口比や GDP 比をとれば、期間を通じてそれほど大きさに差のないデータになるかもしれない。また、対数をとることでも、大きさの違いを緩和できる。

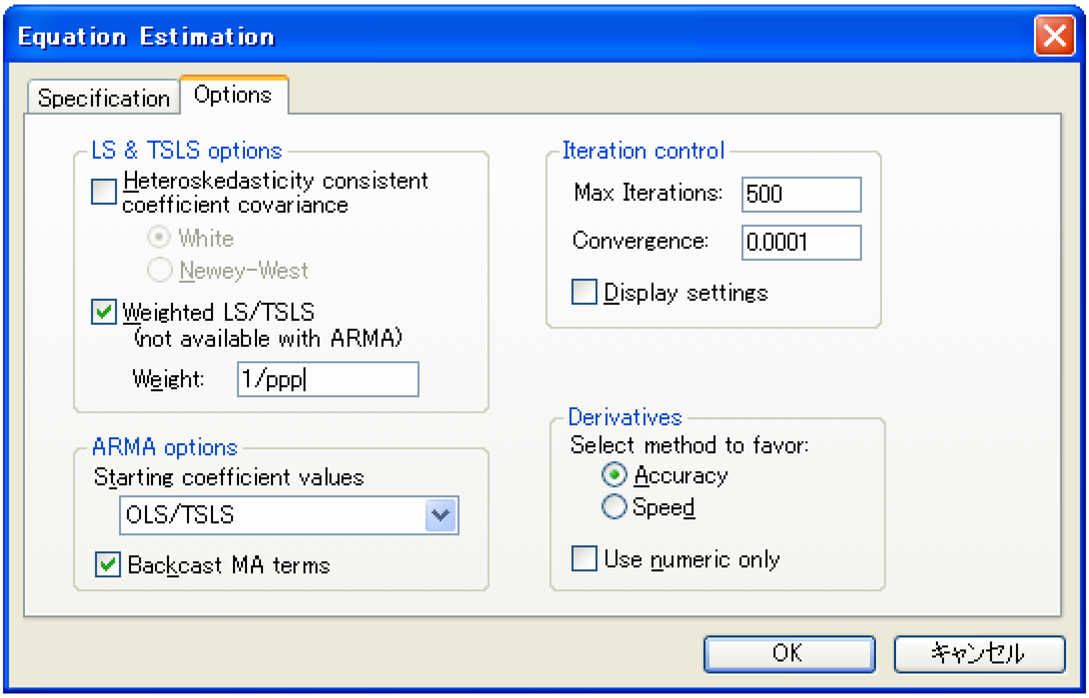
　それでも無理な場合は推計法を変えることになる。誤差の分散が最小二乗法に従わないことを想定して分散が最小になるような推計法を使う。一般化最小二乗法や加重最小二乗法である。

加重最小二乗法(WLS)

　分散が変動する原因の変数がわかっている場合、その変数を使って変数をウエートづけして最小二乗法を行い、加重最小二乗法(Wegihted Least Squares:WLS)と呼ぶ。

加重最小二乗法は、一般化最小二乗法の特殊な場合と考えられる。

　加重最小二乗法は、最少二乗法を推計する画面のオプションの画面で設定することができる。最小二乗法の推計画面では、「specification」側のタグがまず出てくるが、「Options」をクリックすると、次のような画面になる。加重最小二乗法を使うときは、「Weighted LS/TSLS」をチェックし、下のボックスにウエートとなる変数を入力する。ウエートをそのまま入力するのではなく、その逆数を入力するようになっているので注意する必要がある。



為替レートを購買力平価で推計することを考えてみよう。使用したデータは、世界銀行の『世界開発指標 2003』である。2001 年の 32 カ国について対ドルレートを購買力平価を回帰して推計すると、最小二乗法によると次の結果が得られる。中段のカッコ内は標準誤差、下段は t 値である。

為替レートは、韓国のように 1 ドル＝約 1300 ウオンのものもあれば、日本のように 1 ドル＝120 円、またイギリスのように 1 ドル=0.7 ポンドのものまでさまざまな単位であり、購買力平価もほぼ同じ程度の大きさだ。この推計では誤差の分散が不均一である可能性が高い。たとえば、誤差の分散が購買力平価の二乗(PPP^2)に比例していると考えれば、加重最小二乗法は、各変数を PPP で割ったもので推計することである。次の式を最小二乗法で推計することと同じである。

実際の推計では定数項が$¥beta$、説明変数$¥frac{1}{PPP}$の係数が$¥alpha$となる。

加重最小二乗法の推計結果は、次のようになる。$¥alpha$が 0.070、$¥beta$が 1.161 に対応している。

Dependent Variable: EXR
Method: Least Squares
Date: 02/28/06 Time: 12:12
Sample: 1 32
Included observations: 32
Weighting series: 1/PPP

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.069983	0.052416	1.335140	0.1919
PPP	1.161021	0.068928	16.84399	0.0000

Weighted Statistics

R-squared	0.904373	Mean dependent var	1.301586
Adjusted R-squared	0.901186	S.D. dependent var	0.291997
S.E. of regression	0.288380	Akaike info criterion	0.411385
Sum squared resid	2.494889	Schwarz criterion	0.502994
Log likelihood	-4.582167	F-statistic	283.7201
Durbin-Watson stat	1.453269	Prob(F-statistic)	0.000000

Unweighted Statistics

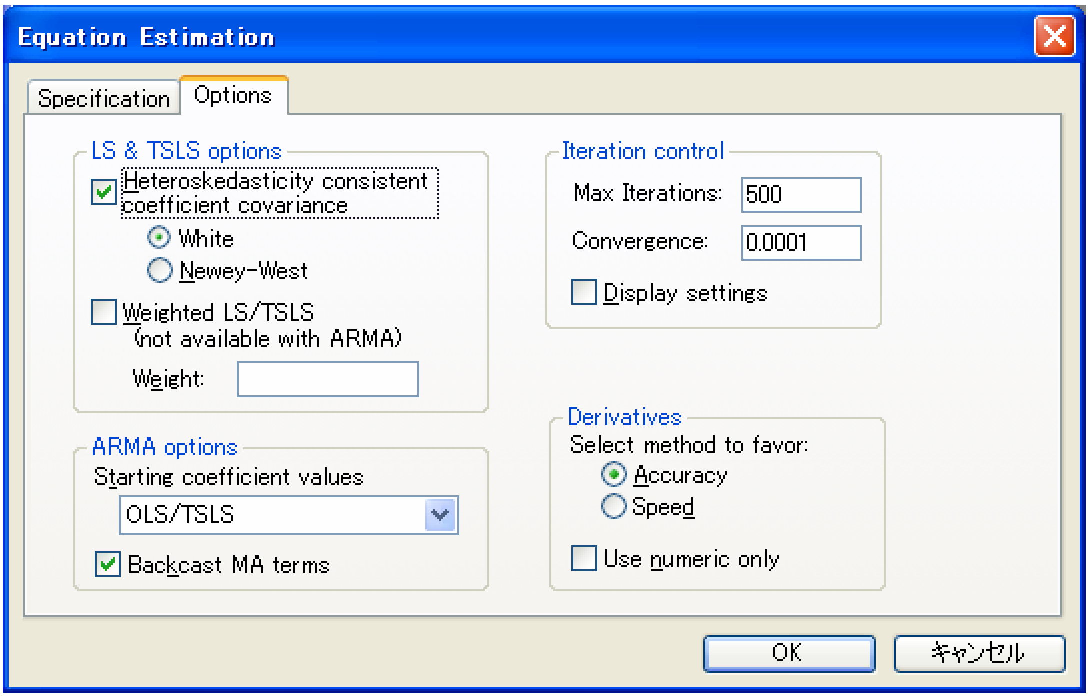
R-squared	0.893388	Mean dependent var	57.06844
Adjusted R-squared	0.889834	S.D. dependent var	230.4335
S.E. of regression	76.48381	Sum squared resid	175493.2
Durbin-Watson stat	0.719354		

ホワイトの分散共分散行列

不均一分散がある場合は、推計された誤差の分散が最小でないことが問題である。そこで、推計された残差を使って、分散共分散行列を計算しなおす方法が提案された（White(1980))。

係数の推定値はそのままにして、標準誤差が通常の最小二乗法で推計したより小さくなり、係数を標準誤差で割った t 値は大きくなる。

加重最小自乗法では、ほかの変数のウエートを使って分散共分散行列を計算したが、ホワイトの方法では、推計残差を使って、次の分散共分散行列を使って標準誤差を推計する。



Dependent Variable: CP95
Method: Least Squares
Date: 02/28/06 Time: 12:18
Sample: 1980Q1 2003Q2
Included observations: 94
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4768.410	1492.399	3.195132	0.0019
GDP95	0.536899	0.003653	146.9552	0.0000

R-squared	0.993307	Mean dependent var	244532.5
Adjusted R-squared	0.993235	S.D. dependent var	42719.09
S.E. of regression	3513.732	Akaike info criterion	19.18779
Sum squared resid	1.14E+09	Schwarz criterion	19.24190
Log likelihood	-899.8262	F-statistic	13654.42
Durbin-Watson stat	0.284721	Prob(F-statistic)	0.000000

ホワイトテスト

ホワイトテストは、不均一分散があるかどうかを検定する方法の一つである。

真のモデルが次のように表せるとする。

誤差が不均一に分散していれば、誤差の分散(e^2)は、ほかの変数と何らかの関係があると考える。そこで、誤差の分散の代理変数として式の推計誤差 e^2 を用い、それが何らかの変数と関連があるかどうかを検定して不均一分散の存在を確かめる。

推計誤差の二乗と相関する変数の候補、つまり不均一分散を引き起こしている原因の変数が判明している場合はその変数と回帰すればよい。しかし通常原因の変数は特定できないため、説明変数や、その 2 乗、さらに説明変数どうしをかけたもの（cross term と呼ぶ）を不均一分散の原因変数とみなす。

誤差の分散が均一なら、これらの係数につてこの仮説検定はラグランジェ乗数検定を行う。定数項を除いた係数の数を s とすると、決定係数をサンプル(n)倍したものが自由度 s のカイ二乗分布に従う。

E Views での操作は次の通りである。

[View]→[Residual Tests]→[White Heteroskedasticity Test(no cross term)]

ホワイトテストによって、不均一分散の有無を検定してみよう。誤差項の二乗に、説明変数と、説明変数の二乗を回帰させる。検定統計量は決定係数にサンプル数（32）をかけたもので、0.528736*32＝16.91955 である。このときの p 値は 0.000212 と小さく、1%有意水準で「定数項を除くすべての係数＝ゼロ」という帰無仮説が棄却できる。つまり、誤差が不均一分散であることがわかる。

White Heteroskedasticity Test:

F-statistic	7.617257	Prob. F(2,91)	0.000874
Obs*R-squared	13.48003	Prob. Chi-Square(2)	0.001183

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 02/28/06 Time: 12:26
Sample: 1980Q1 2003Q2
Included observations: 94

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.57E+08	61106823	-2.576636	0.0116
GDP95	750.7042	292.5722	2.565877	0.0119
GDP95^2	-0.000806	0.000340	-2.371497	0.0198

R-squared	0.143405	Mean dependent var	12083627

Adjusted R-squared	0.124578	S.D. dependent var	15372268
S.E. of regression	14382904	Akaike info criterion	35.83237
Sum squared resid	1.88E+16	Schwarz criterion	35.91354
Log likelihood	-1681.122	F-statistic	7.617257
Durbin-Watson stat	0.630692	Prob(F-statistic)	0.000874

Goldfeld-Quandt テスト

ある変数を大きい順に並べる。その間の幾つかの変数（サンプルの真中３分の１など）を除き、小さい変数だけについて回帰した時の誤差の二乗和を RSS_2、大きい変数について回帰したときの誤差の二乗和を RSS_1 とし、次の統計値を調べる。

$$R = \frac{RSS_2}{RSS_1}$$

この統計量は F 分布をする。

誤差項の系列相関

系列相関とは、ある系列の当期の値が過去の期の値と相関しているものである。不均一分散の一種で、「誤差の分散が時間を通じて一定」という仮定を満たしていない。このため、推計した係数の分散が最小にならず、そこから計算された t 値なども過大になってしまう。残差のグラフを描いてみると、ある時期には上向きの数値が続き、ある時期には下向きの数値が続くことがある。こうした場合は、想定したモデルの誤差項が前の期の誤差項の影響を受けていることを示している。

誤差項に系列相関があるかどうかはダービン・ワトソン比で調べる(¥ref{dw}参照)。ダービン・ワトソン比は当期の誤差と 1 期前の誤差の相関を調べたものだから、系列相関があると 2 から離れているはずである。

解決法としては、誤差項に明示的に系列相関を示す式を作って、推計することが考えられる。最後の式は、係数が入り組んでおり、通常の最小二乗法では推計できない。コクランオーカット法、最尤法などで推定することになる。

系列相関のある場合は不均一分散の一種であることは、誤差項の分散共分散行列が次のように表されることでわかる。$¥rho$は 1 期前の誤差との相関係数である。

ニューイ・ウエストの分散共分散行列

ホワイトの分散共分散行列は、不均一分散を考慮して分散共分散行列を計算し直すものだったが、ニューイ・ウエスト（Newey and West(1987))は不均一分散ともに、誤差の自己相関をも考慮した分散共分散行列の計算法を提案した。

分散不均一（heteroscedasticity)と自己相関（autocorrelation）に対応できるので、HAC 分散共分散行列と呼ばれる。

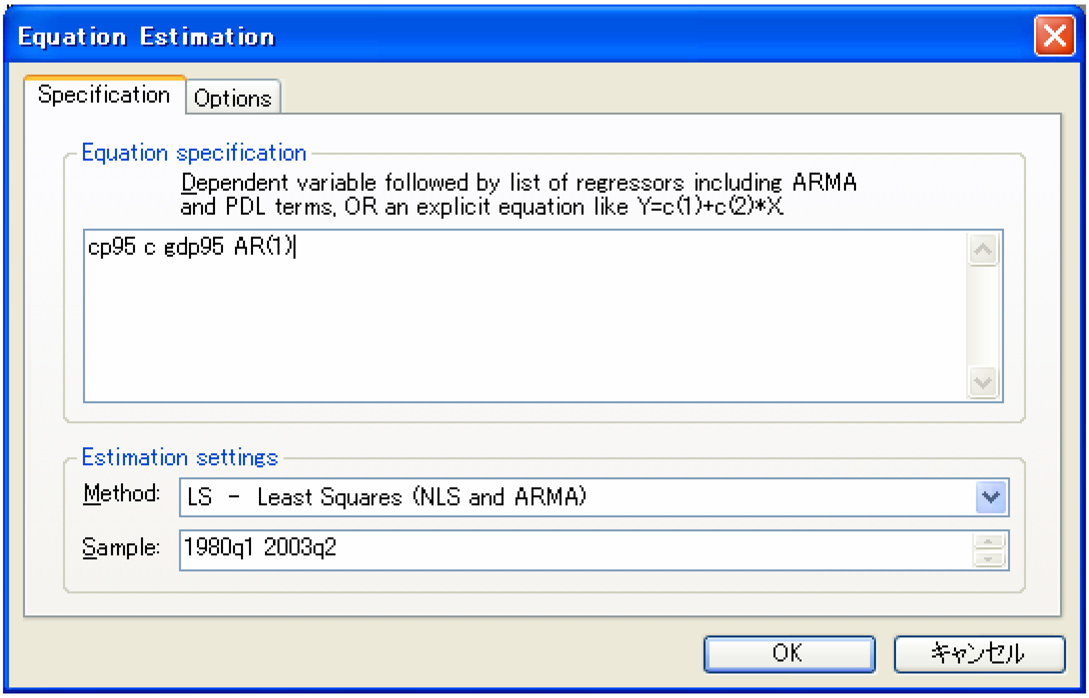
自己相関のラグ q は、サンプル数を T として次の計算値の小数点以下を切り捨てた整数を使う。

ニューイ・ウエストの分散共分散行列は次のように表される。

¥[　ニューイウエストの分散共分散行列＝¥mathbf{ (X'X)^{-1}}　S ¥mathbf{(X'X)^{-1} }
計算例

簡単な計算例を使って、推計値がどのように変わるかを見てみよう。次のような消費関数を推計する。被説明変数は実質民間最終消費支出(CP95)、説明変数は実質 GDP(GDP95)である。推計期間は 1980 年から 2003 年 4－6 月期までである。最小二乗法で推計すると次の結果が得られる。中段のカッコ内は標準誤差、下段は t 値である。

ダービンワトソン比が 0.285 と 2 から大きく離れており、残差が系列相関していることがわかる。誤差の系列相関は不均一分散の一種であるので不偏性が成り立たない。係数の真の標準誤差は計算結果（0.004595）より大きいはずで、116.85 という t 値も過大評価されている。
こうしたケースではまず、説明変数を増やして残差の系列相関を無くすことを検討するのが常道だが、ここでは
残差に１次の系列相関を仮定して推計してみる。統計ソフト EVIEWS では説明変数の後に「AR(1)」という変数を加えることで推計できる。



両推計の違いは誤差の動きの違いに端的に表れる（¥ref{gosaar}参照）。
最小二乗法で推計した場合は、誤差どうしの相関が強く、誤差がプラスになるとしばらくプラスの値が続き、マイナスになるとしばらくマイナスの値が続くことがわかる。
一方系列相関を除去した場合の誤差の動きはランダムな動きに近づいている。ダービンワトソン比は 2.127 と 2 に近い。この推計では
推計結果によると、次のような式にしたがっている。

Dependent Variable: CP95
Method: Least Squares
Date: 02/28/06 Time: 12:33
Sample (adjusted): 1980Q2 2003Q2
Included observations: 93 after adjustments
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3159.428	9122.334	-0.346340	0.7299
GDP95	0.553537	0.019188	28.84797	0.0000
AR(1)	0.866923	0.054954	15.77539	0.0000

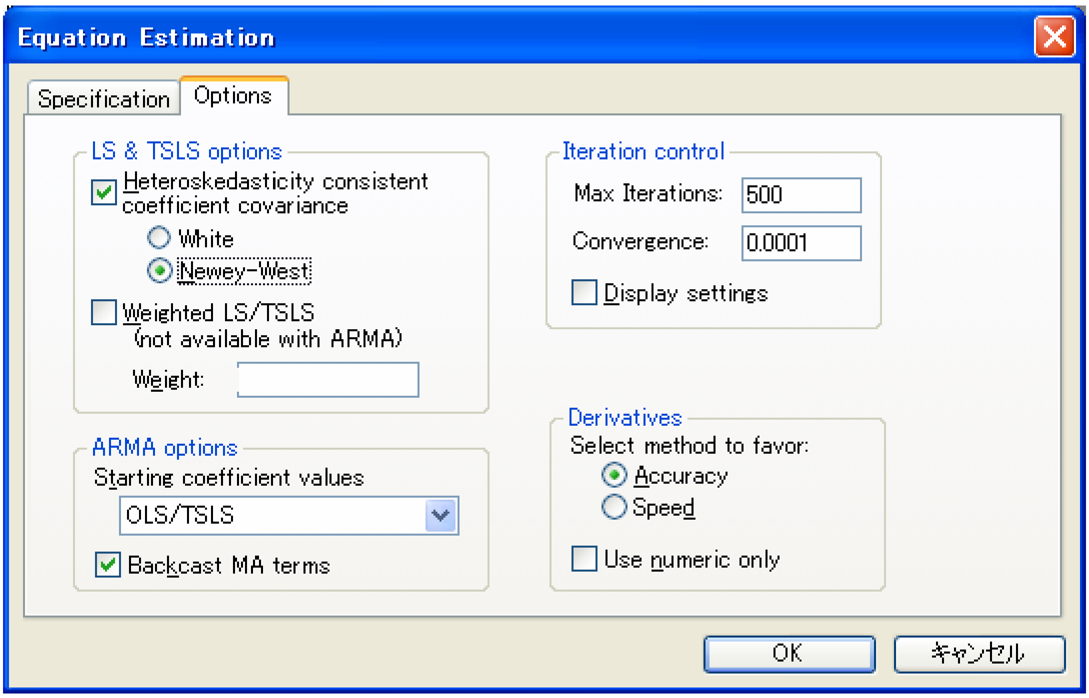
R-squared	0.998198	Mean dependent var	245305.9
Adjusted R-squared	0.998158	S.D. dependent var	42283.82
S.E. of regression	1814.879	Akaike info criterion	17.87715
Sum squared resid	2.96E+08	Schwarz criterion	17.95885
Log likelihood	-828.2875	F-statistic	24924.57

Durbin-Watson stat	2.127230	Prob(F-statistic)	0.000000
Inverted AR Roots	.87		

GDP95 にかかる係数の大きさは 0.54 と 0.55 とそれほど違いはないが、標準誤差や t 値の動きは大きく異なっており、
通常に推計した場合の t 値は 116.85 だかが、系列相関を除去した場合は 28.85 である。このケースでは両方とも有意
であるが、系列相関を除去しない場合は係数が有意だが、除去後は有意とならない場合もある。推計法としては、後者を使うほうが望ましいことがわかる。

次に誤差の自己相関を明示的にモデルに組み込まず、ニューイウエストの分散共分散行列だけを使った場合である。

係数は、最小二乗法と同じであるが、標準誤差が通常の最小二乗法より大きくなり、t 値が小さくなることがわかる。
コクラン・オーカット法タイプの場合では自己相関を 1 期前しか想定していないが、この推計ではサンプル数が 94 なので



説明変数が確率変数である場合

最小二乗法の仮定の一つは「説明変数が確率変数ではない」である。説明変数x_tは他

の変数から影響を受けない「地に足のついた」データであるという仮定だ。
　しかし、経済データは概してほかの変数に影響され、ほかの経済変数から独立して成り立つ変数は稀にしかない。消費は所得の影響を受け、投資は金利や企業収益の影響を受ける。輸出は海外の需要動向の影響を受け、輸入は国内の需要や輸入価格に影響を受ける。公共投資は政府が決定するため、ほかの変数の影響を受けないとの見方もあるが、ＧＤＰや雇用情勢など経済状況で影響される。日本経済にとって、独立して決まるのは原油価格くらいかもしれない。

何が問題か
「説明変数が確率変数である」という仮定が崩れると何が問題になるのだろうか。
最小二乗法の係数の期待値は次のように表される（¥eqref{huhen2}式参照、$¥tilde x_t =x_t- ¥bar x$ に変換)。

x_tが確率変数でなければ、$E(x_t)=x_t$と処理でき、$E(u_t)=0$なので、$E(b)=¥beta$となる。
x_tが確率変数のときは、x_tには何らかの期待値が入り右辺の第 2 項はゼロとならない。つまり、
係数の推計値 b は、真の値と異なる値を推計することになる。

¥[E(b) ¥neq ¥beta ¥]

x_tが確率変数であれば、最小二乗法の望ましい性質の一つである不偏性(¥ref{huhensei}参照）が成り立たないこととが問題である。

次にサンプル数が増えると、真の値に近づくかどうか（一致性）を調べてみよう。
¥eqref{gosa1}式の右辺第 2 項を標本数nで割る。分子はx_tと誤差u_tの標本共分散($s_{xu})$)、分母はx_tの標本分散(s^2_x)を表す。

サンプルの分散や共分散はサンプル数が増えれば母集団の分散($¥delta_{xu}$)、共分散($¥delta^2_x$)に収束する。
つまり、次の式が成り立つ（確率極限については¥ref{plim}参照)。

右辺第 2 項の　x_t と　u_tに相関がなければ推計値 b は確率的に　$¥beta$ に収束

し、一致性を満たす。
x_t と　u_t　に相関があれば不偏性とともに一致性も満たさないことになる。

さまざまなケース
誤差項と説明変数に相関ない場合は、説明変数が確率変数であることはそれほど問題ではなく、通常の
経済データを使って推計しても一致性はある。
問題なのは誤差項と説明変数に相関がある場合である。
実際の推計ではどのようなケースが問題となるのかを見ていこう。説明変数が
確率変数とならない代表的な例は次の通りである。

こうした問題を解決するには操作変数法という推計法が使用される。

¥subsection{被説明変数のラグが入っている場合}

説明変数に被説明変数のラグが入っている場合のモデルは次のようなものである。

説明変数であるy_{t-1}は、y_tという確率変数の一期前の値なので、確率変数である。ただ、y_{t-1}は当期時点では確定した値なので、当期の$¥varepsilon_t$ との相関はない。
この場合は、一致性は保たれる。つまりサンプルが多い場合は深刻な問題を生むものではない。

観測誤差がある場合
統計には誤差がつきものである。GDP 統計も毎四半期のように改定される。つまり、本当に推計したいのは確報値のx^*_tであるのに、誤差のある速報値x_tを使って推計してしまう場合がある。このとき、速報値と確報値の間には次の関係があるとする。v_tは誤差項とする。

データとしてxを使うと、いうことは次の式を推計することだ。

しかし本当の式は、次の式である。

係数が「薄められれる」 と呼ぶ。

連立方程式
　たとえば、GDP を説明変数に使う場合、ほかの変数からの影響を受けない外生変数と考えるのには無理がある。推計誤差とx_tに相関があるということになる。
GDP をY_t,消費をC_t,投資をI_tとした次のようなモデルを考える。

¥[C_t=a_0+a_1 Y_t+u_t　¥]
¥[Y_t=C_t+I_t　 ¥]

2 式を使ってY_tについて解くと次の式になる。

¥[Y_t= ¥frac{a_0}{1-a_1}+¥frac{I_t}{1-a_1}+¥frac{u_t}{1-a_1} ¥]

これはu_tがY_tに影響していることを示しており、Y_tがu_tから独立ではない確率変数であることを示している。

¥subsection{操作変数法　}　　¥¥　¥index{そうさへんすうほう@操作変数法}
　x_tと誤差項との間に相関がある場合の解決法するには、操作変数法を使う。
推計したいのは次式でXが確率変数の場合である。

¥[y_t=a+b x_t+ e_t　¥]

推計式の左辺にある変数は被説明変数と呼ばれ、右辺の変数は説明変数と呼ばれる。これに加えて、新たに「操作変数(Instrment Variable)」という変数を導入してみよう。
操作変数と聞いただけでは何を「操作」するのかわからない。省略せずに言えば「説明変数X_tを操作する変数」である。
　最小二乗法ではx_tが非確率変数であることを仮定している。つまり、ほかの変数が変化しても動じない、「地に足の着いた」データを想定しているが実際には確率的に変動する可能性がある。

　変数が確率変数で誤差と相関している場合、最小二乗法を適用しても不偏性も一致性もない。これを解決する方法として、Xの性質を変えてしまおうというのが操作変数法だ。ぐらぐらしているXを地に足のついたデータに変換する。

操作変数の考え方を簡単に説明すると次の通り。
確率変数を地に足の着いたものにするため、Xに操作変数Zを回帰させる。

操作変数のZで、Xの足場を固めたもの(¥eqref{eq:iv1}式の推計値)を
$¥hat X$とする。

¥[¥hat X=c+dZ　¥]

次に、この推計値をYに回帰して係数を求める。

¥[y=a' + b' ¥hat X+e　¥]

こうして求められた係数a'、b'は一致性を持つ。
これが操作変数法の手順である。2 つの推計を行っているが、
計算上はこれを一括して行うことがで、それで得られる係数は次の式で表される。

¥[b_{IV}=(Z'X)^{-1}(Z'Y)　¥]

¥eqref{eq:iv1}式を用いる以上、なるべくXとZの相関は高いほうが良い。
一方、ZがYと相関があっては、足場を固める変数の意味がなくなってしまう。
つまり、操作変数ZはなるべくXと相関が高く、Yと相関のないものを選ぶ必要がある。

操作変数の候補
　操作変数の候補は次のものである。
まず、定数項である。これは最小二乗法を使う場合には通常必要となりうう。
説明変数の 1 期前は推計時には決まった変数であり、概してXとの相関は高い。
また、複数本のモデルが想定されている場合は、モデルから考えられる外生変数も候補となる。

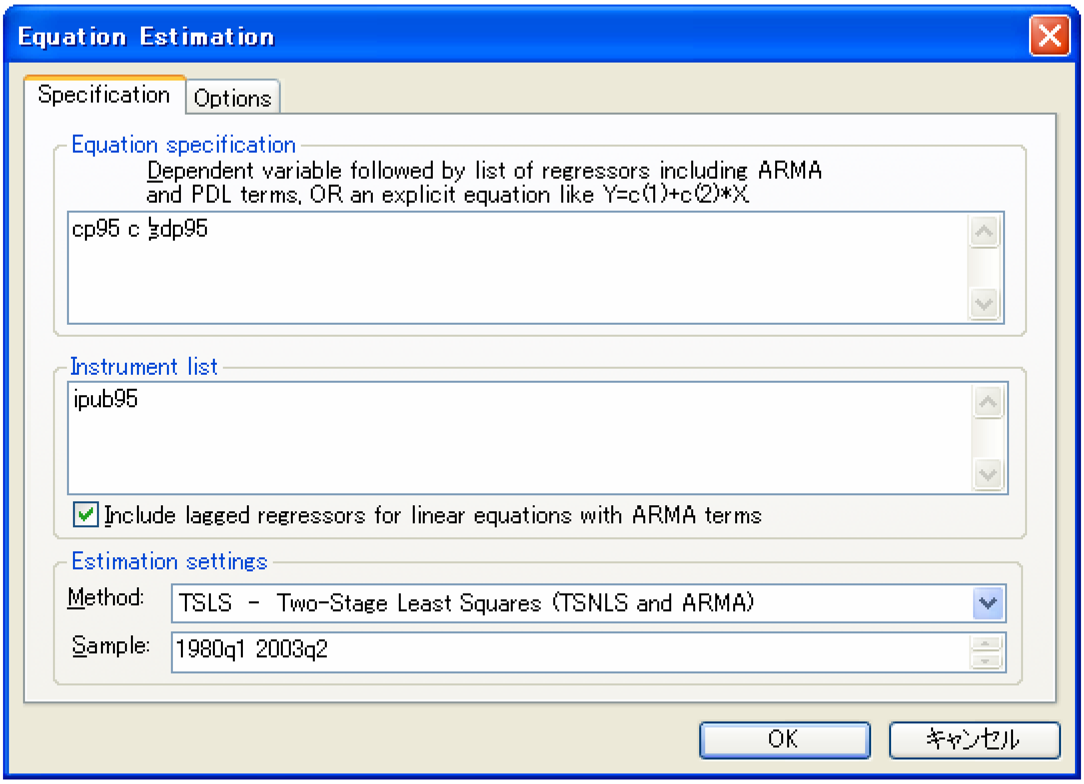
操作変数の数は識別条件によって決まる。

識別可能になるためには、定数項を除く説明変数以上の操作変数が必要である。

２段階最小二乗法
複数の方程式システムの場合。操作変数として、外生変数すべてを使った場合。
手順は操作変数法と同じである。

具体例
　最小二乗法の仮定では、実質 GDP と誤差項の間に相関がない。しかし、実質 GDP は
実質 GDP がその構成項目である消費の影響を受けていることは避けられない。
そこで、操作変数として実質公的固定資本形成（IPUB95)を使うことにする。推計結果は
以下の通り。
　係数が、0.537 から 0.553 へと変化した。操作変数法を通常の最小二乗法で推計する場合
は
次の 2 段階の手順を踏む。まず説明変数を操作変数で回帰する。

上式の推計値（$¥hat {GDP95_t} =131848.1+9.60004IPUB95_t$)を使って、消費を関数を
推計すると、操作変数法と同じ結果が得られる。



Dependent Variable: CP95
Method: Two-Stage Least Squares
Date: 02/28/06 Time: 12:34
Sample: 1980Q1 2003Q2
Included observations: 94
Instrument list: IPUB95

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2504.369	2831.189	-0.884564	0.3787
GDP95	0.553185	0.006281	88.07956	0.0000

R-squared	0.992393	Mean dependent var	244532.5
Adjusted R-squared	0.992311	S.D. dependent var	42719.09
S.E. of regression	3745.972	Sum squared resid	1.29E+09
Durbin-Watson stat	0.247790	Second-stage SSR	6.09E+10

説明変数間に相関がある場合（多重共線性）

　最小二乗法の仮定がすべて満たされていても問題が生じる場合がある。説明変数に相関がある場合で、多重共線性（マルチコリアニティー）と呼ばれている。

　説明変数が似たものが入っていると、いずれかの変数がかなりおかしな値になる。サンプル数の増減に影響を受けやすくなる。

　簡単な例を示してみよう。賃金を物価と実質ＧＤＰで回帰させてみる。

まとめ

 最小二乗法で推計することが多いのは、不偏性や有効性など推定量として望ましい性質を持っているためだ。しかし、それは多くの仮定のもとに成立するものである。

中でも問題なのは、「説明変数は確率変数でない」という仮定である。ただ、単に独立変数でなくても誤差項と相関がない場合はそれほど問題とはならない。

被説明変数のラグが説明変数にある場合、その変数は当然確率変数ではない。しかし、この場合でも一致性はあるので、

ただ、この仮定が満たされなくても、係数の一致性はあるので、標本数が多ければ推計上の問題はない。

ただ、説明変数に観測誤差がある場合や同時方程式モデルのうちの一本を推計する場合は問題である。

いずれも誤差と説明変数に相関が生じてしまう。この状態で推計値を推計すると、

不偏性もなく、有効性もない推定量となる。つまり、推計された係数は、真の係数からずれて推計されるうえ、係数の有意性を見るための

t 値は過大に算出される。こうした問題を解決するには、操作変数を使うのが望ましい。