

ネコでもわかる EViews 入門

一橋大学国際・公共政策大学院

別所俊一郎

1. はじめに

EViews は、とくに官公庁で広く使われているといわれている計量経済パッケージの 1 つです。製造元のサイト (<http://www.eviews.com/>) を訪れてみると分かるように、EViews の特徴は、「modern windowing and object-based techniques」にあります。代表的な時系列解析ソフトウェアとして知られる TSP¹から分化して開発された MicroTSP を前身とする EViews は、オブジェクト指向の操作環境を実現しており、比較的「とっつきやすい」計量経済ソフトウェアとなっています。プログラムを書いたりコマンドラインを手入力したりする必要性をおさえ、ポインティングデバイスを用いてツールバー経由で操作が可能なユーザインターフェイスは、いまでこそ Stata²にも導入されていますが、EViews で先駆的に取り入れられたものです。

EViews は TSP から分化して開発されてきたという経緯もあつてか、時系列解析にその強みを持っており、時系列データを用いる分析に便利なような仕様になっています。また、時系列のグラフ機能も豊富なようです。もちろん、それほど複雑でない分析であればクロスセクションやパネルデータにも十分に対応しています³。基本的には EViews は対話型での使用が想定されているように思われます⁴。

計量経済分析には、多分に「慣れ」が必要であるような気がします。某民間調査機関の「回帰千本ノック」とまではいかなくとも、いろいろな推定手法を知り、実際にさまざまなデータを用いて行なってみることで理解が深まっていくかもしれません⁵。

¹ Time Series Processor の略称らしい。ちなみに RATS は Regression Analysis of Time Series の略称なんだとか。

² <http://www.stata.com/>

³ 大規模な個票等を扱う場合には、個人的には Stata のほうが便利ではないかと思います。

⁴ 現状の TSP のように、処理全体のプログラムを書いて一括して処理させる（バッチ方式）ことを前提としているソフトウェアもあります。

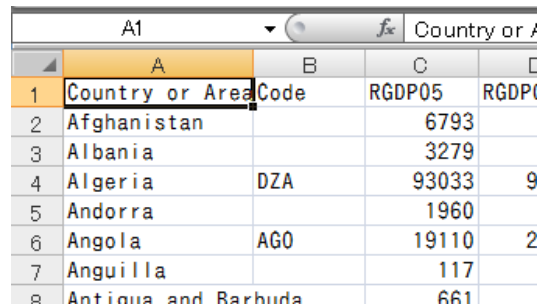
⁵ かつて経済研究所に在籍しておられた小西葉子先生たちのマニュアルは優れたものです。なくなるまえに取っておくといいと思います。 <http://www.ier.hit-u.ac.jp/~konishi/text.html>

2. データの読み込み

どのような計量経済パッケージを使うにしても、最もつまづきやすいのが「データの読み込み」です。いかにユーザフレンドリーなインターフェイスをしていても、いかに高度な計量手法を簡単なコマンドで実行してくれるとしても、データを読み込んでくれないと話が進みませんが、ここでうまくいかないでイライラが募る人はけっこういます。ほら（経験者談）。

2.1. データ形式と変数名

EViews は、それ自身によるデータ全体のハンドリング（合体やマッチング）が得意ではなさそうなので、基本的には表計算ソフト上でデータを整理しておく必要があります。時系列データを例にとると、時系列に沿って縦にデータを並べていっても（by observation）、横に並べていっても（by variable）でもどちらでもかまいませんが、EViews での分析を前提にするなら、by observation で揃えたほうがよいです。ここで、データを構築する際に注意するポイントがいくつかあります。



	A	B	C	D
1	Country or Area Code		RGDP05	RGDP05
2	Afghanistan		6793	
3	Albania		3279	
4	Algeria	DZA	93033	9
5	Andorra		1960	
6	Angola	AGO	19110	2
7	Anguilla		117	
8	Antigua and Barbuda		661	

こんなかんじ。1 行目の変数名、2 行目より下がデータ。
A 列と B 列に文字データが入ってますが気にしない気にしない。

変数名以外のデータは小数表記にしましょう⁶。Eviews も文字データを扱ってくれるようですが、文字データの処理が得意ではなさそうなので、数値情報に変換しておいたほうが無難でしょう。経済主体やその他の文字情報で表される変数を扱うときには、代理のダミー変数を作るか、識別番号をデータとする ID 変数を作っておきましょう。その場合、どの数字がどれを表すかを忘れないように。都道府県に 1 から 47 までの番号を振ってみたものの、何番がどの県かわからなくなってしまって困る、ということもあ

⁶ 桁区切り「,」も念のため外しておいたほうがいいかもしれません。

るかもしれません⁷.

一番上の行, EViews 用の変数名をつけましょう. 変数名の長さに特段の制限はありませんが, だからといってやたら長くするととにかくと不便なので, 分かる範囲で短くしましょう. 古いバージョンでは 16 文字以内らしいので, おおむね 10 文字以内に収めておくのが望ましいでしょう. 新たな変数を定義したりする場合に, 元の変数の名前を組み合わせたたりすることが多いからです. また, 変数名に用いることができるのは半角英数字と半角アンダーバー (_) といったところですが, 大文字と小文字は区別しません (だって DOS なんかもーん). 数字を変数名の先頭に持ってくることはできません. 半角記号の一部も変数名に使うことができますが, EViews 特有の使い方をする場合があるので, 特に必要がなければ英数字とアンダーバーだけで命名しておくほうが無難でしょう. 計量経済ソフトウェアによくあることですが, いくつかの名前にはすでに EViews 自身が特殊な役割を与えているので, こちらで変数名として使うことはできません. 以下のとおりです.

```
ABS ACOS AR ASIN C CON CNORM COEF COS D
DLOG DNORM ELSE ENDIF EXP LOG LOGIT LPT1 LPT2 MA
NA NRND PDL RESID RND SAR SIN SQR THEN
```

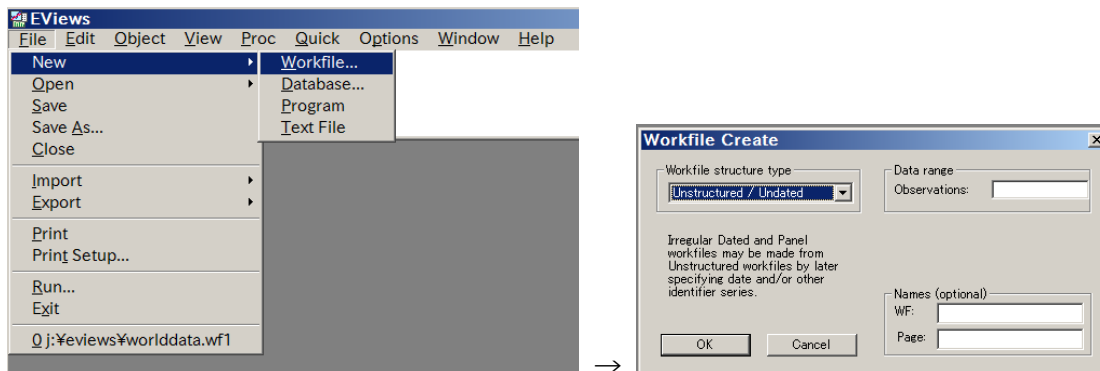
これ以外は基本的には変数名にしてもいい (構文で判断されるから) のですが, コマンド名と同じ変数名は自分のためにも避けたほうが無難です. 変数名をつけたあとは, それぞれの変数名とその定義の対応関係がわかるようなメモを残しておきましょう. 慣れないうちは, EViews 読み込み用のファイルを作り, 1 行目に変数名, 2 行目以下にデータを格納して, データの定義などは別のファイルに記述しておくといいでしょう.

2.2. Workfile

EViews では, 読み込んだデータなどを EViews 特有のデータ形式に変換して扱い, 保存します. そのときに利用されるのが workfile と呼ばれるもので, 作成したデータ, 推計結果, グラフ等の全作業結果を格納するフォルダのようなものです. 分析を始めるときには新規に workfile を作成します. 保存された workfile を呼び出してそこに格納されているデータを用いて分析の続きを行うこともできます. ここでいうデータの読み込みとは, 新規に workfile を作成し, そこに手許のデータを格納する作業までを言います.

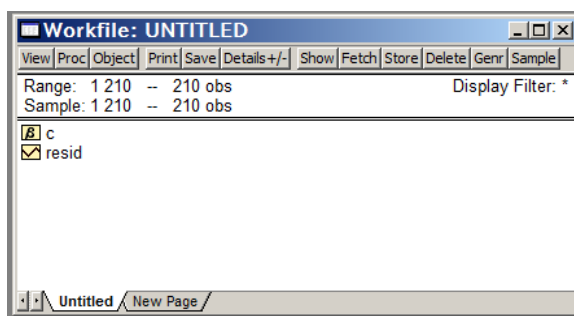
⁷ 都道府県くらいなら都道府県コードをおぼろげに覚えていることもあります (A 大学の N 先生みたいに覚えている人もいるかもしれませんが), 市町村や国のデータだとそうはいきません. たぶん.

EViews を起動し、メインメニューの File からドロップダウンメニューの New → Workfile を選ぶと、Workfile Create ダイアログが表示されます。



こんなかんじ。ここでは unstructured/undated を選んでみました

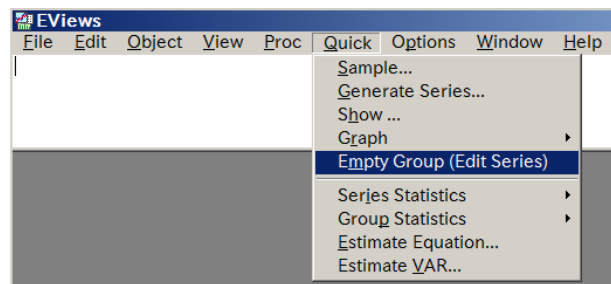
まずデータの周期 (frequency) を聞かれます。構造なし (Unstructured/Undated), 時系列構造 (Dated – regular frequency), バランスド・パネル (Balanced panel) からの選択ですが、クロスセクションデータあるいはアンバランスド・パネルデータの場合は構造なしを選び、右のボックス (Observations) に、観測値数を入力します。時系列データの場合は右のボックスからデータの周期 (年次, 半期, 四半期, 月次, 週次, 日次) を選び、4 桁の西暦表現で始期と終期を入力し、年と四半期・月はコロン (:) で区切ります。OK をクリックするとダイアログが閉じ、c と resid だけが表示される空の workfile が作成されます。メインメニューの File からドロップダウンメニューの Save as を選ぶと名前をつけて保存することができますが、ファイルのパスに日本語が混じっているとエラーが出るので気をつけましょう。



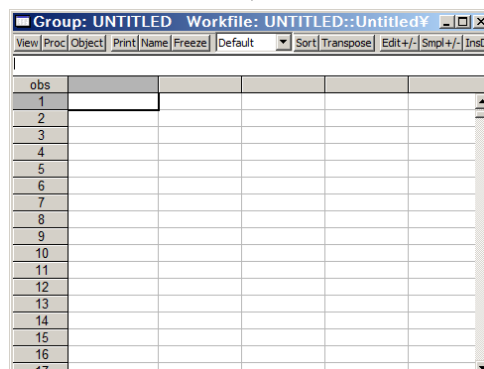
からっぽの workfile。上部にたくさんのボタンが並んでいます。

さて、データを格納する workfile を用意できたら、ここへデータを取り込みます。表計算ソフトの MS-Excel を用いてデータの整理をすることが多いと思うので、ここからデータを読み込んでみましょう。やり方はいくつかあります。

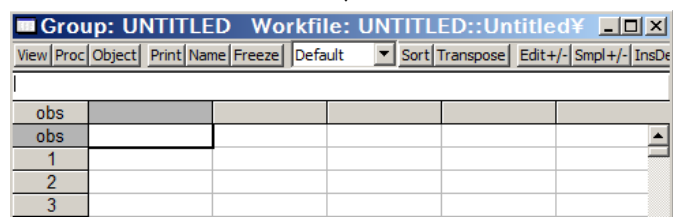
Copy and paste による方法(1)： MS-Excel で変数名とデータの範囲を選択してコピーしておきます。続けて、メインメニューの Quick からドロップダウンメニューの Empty Group と選ぶと、スプレッドシートが出現するので、obs が 2 個見えるようにしておいてから、下の obs のすぐ右隣のセルをアクティブにしてから貼り付けてやります。貼り付けたウィンドウを閉じると、ワークファイルウィンドウに、c と resid 以外に新しい系列 (series) が加わっていることが確認できるでしょう。このとき、変数名がちゃんと認識されていることを確認しましょう。



ここをクリックするとウィンドウが開きます



最初はこんなふうに出てしまうので、右のカーソルを少し上に動かして



こうやっておいて左上隅をクリックして貼り付けます。すると。



obs	ALPHA01	CODE	RGDP05	RGDP06	POP05
1	Afghanistan		6793.000	7295.000	25
2	Albania		3279.000	3460.000	31
3	Algeria	DZA	93033.00	94893.00	32
4	Andorra		1960.000	2093.000	73
5	Ananla	AGO	19110.00	22658.00	16

こんなふうに読み込まれます。上の obs の右側に変数名が入っていれば成功。
このウィンドウは閉じてしまいましょう。Delete してよいです。

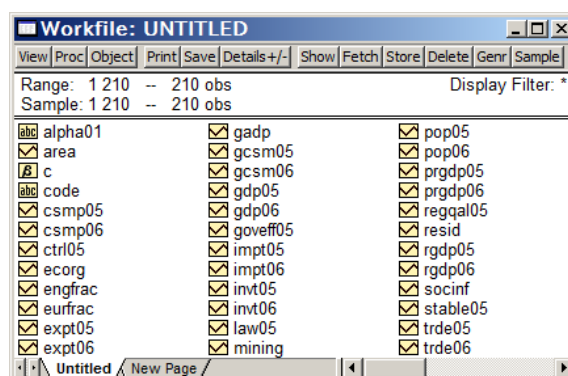
Copy and paste による方法(2): メインメニューの Objects からドロップダウンメニューの New Object を選ぶと、New Object ウィンドウが開きます。ここで Type of object として「Series」を選んで Name for object を入力すると、Series ウィンドウが開くので、Edit 状態を確認します。MS-Excel を開いて入力したデータ（変数名は含まない）をコピーし、Series ウィンドウに貼り付ければデータを追加したことになります。ワークファイルウィンドウで系列が追加されていることを確認しましょう。このとき、すでに何らかのデータを読み込んでいるときには、観測値番号が一致しているかどうかにも注意しましょう。自信がない場合には、MS-Excel などデータを整理して一括して読み込むほうが賢明です。

Import する方法(1): あらかじめ MS-Excel でデータファイルを開き、データ系列（変数名含まない）の左上のセルの位置（A2 とか）を覚え、変数名のリストをコピーしておきましょう。EViews のメインメニュー File からドロップダウンメニューの Import → Read Text-Excel-Lotus を選びます。Open ウィンドウが開くので、読み込みたい MS-Excel ファイルを選択します。Upper-left データで覚えておいた左上のセルの位置（A2 とか）を入力し、Name for series or Number if named in file の欄にコピーしておいた変数名リストを貼り付けます。これで読み込みが行われます。この方法では、CSV ファイルやテキストファイルからもデータの読み込みを行うことができます。

Import する方法(2): メインメニューの File → Open → Foreign Data as Workfile とたどって、MS-Excel ファイルを開く方法もあります。このとき、開かれる MS-Excel ファイルのシート名に日本語など（次小節参照）が混じっているとエラーが出るかもしれません。

いずれにしても、ワークファイルウィンドウを開いて変数名が表示されているかどうかを確認し、それぞれの変数名をダブルクリックして系列を表示させ、各系列がちゃんと取り込まれているかどうかを確認しましょう。変数名のコピー・ペーストがうまくいっ

ていないなどの理由で、数字が取り込まれているもののなんだか頓珍漢なことになったりしていないでしょうか。データの取り込みがうまくいったら、さっそくメインメニューの File からドロップダウンメニューの Save as を選んで保存しておきましょう。保存するときには、単精度（single precision）か倍精度（double precision）かを尋ねられますが、倍精度にしておきましょう。



うまくいっていると、Workfile ウィンドウに変数名がずらずら並びます。

2.3. データの読み込み、作業時の注意点

EViews は見たとおりの英語のアプリケーションで、日本語を認識してくれません。なので、保存するワークファイルの名前はすべてアルファベットと数字のみとしましょう。また、データの読み込み、保存、作業などでは以下のような点に気をつけましょう。

- ・ワークファイルを保存したフォルダのパスのなかにアルファベットと数字以外の文字（漢字やひらがな、全角文字、@、#などの記号）が入っていると、ワークファイルを読み込むことができません。また、このときにはファイルの import もできません（よくない例：D:¥一橋¥計量経済¥EViews）。
- ・ファイルを import してデータを読み込むとき、読み込まれるファイルが MS-Excelなどで開かれている場合には読み込みが行われません。
- ・ファイルを import してデータを読み込むとき、読み込まれるファイル名にアルファベットと数字以外の文字（漢字やひらがな、全角文字、@、#などの記号）が入っていると読み込めません。

3. 変数の変形

変数を変形するという手続きはほとんどの実証分析で行われます。対数変換をしたり、2次関数を作ったり、交差項を作ったり、時系列データなら差分をとったり変化率をと

ったりします。これらの手続きは、新たな変数を定義することによって行いますが、EViews の場合には、新たな変数を定義しなくても回帰分析自体は行うことができます。

新たな変数の定義には、GENR コマンド (SERIES コマンド) を使います。メニューバーのすぐ下にあるコマンドウィンドウにコマンドラインを入力することもできます。たとえば2乗項を作る場合、あるいは対数変換する場合、

```
GENR X2 = X * X
```

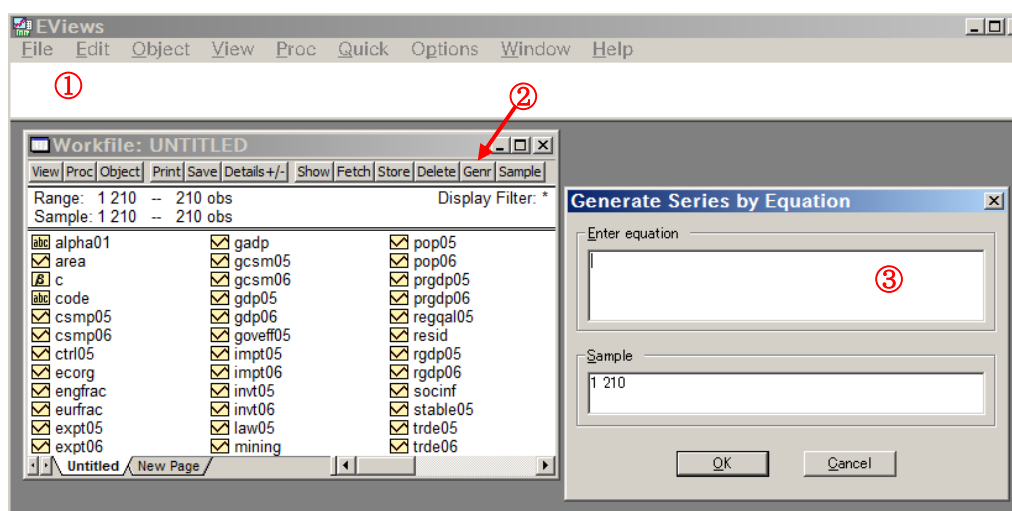
```
GENR LX = LOG(X)
```

時系列データで階差をとる場合と変化率をとる場合には

```
GENR DX = D(X)
```

```
GENR PX = @PCH(X)
```

コマンドウィンドウにコマンドラインを入力しなくても、ワークファイルウィンドウのメニューバーの GENR をクリックし、出てきたウィンドウに上記コマンドラインの GENR よりあとの部分だけ入力しても同じことです。四則計算はもちろん、上述の対数変換、累乗 (^) や指数関数 (exp) が使えます。階差・変化率以外にも、時系列データに特有な変換もいくつか用意されています。



上の白い部分 (①) に入力するか、
②をクリックしてウィンドウ (③) の上のボックスに入力しましょう

ダミー変数を作るには、以下のようなコマンド (ワークファイルウィンドウのメニューバーの GENR をクリックして下記の GENR よりあとの部分だけ) を入力します。たとえば、変数 avginc が 14 より小さい観測値に 1 を与えるようなダミー変数 low は

```
GENR LOW = AVGINC < 14
```

のように作ります。書式としては、「新しい変数」 = 「論理式」となります。したがっ

て、変数 PREF において東京都が 13 という値を取るとして、東京都ダミーを作りたい場合には、

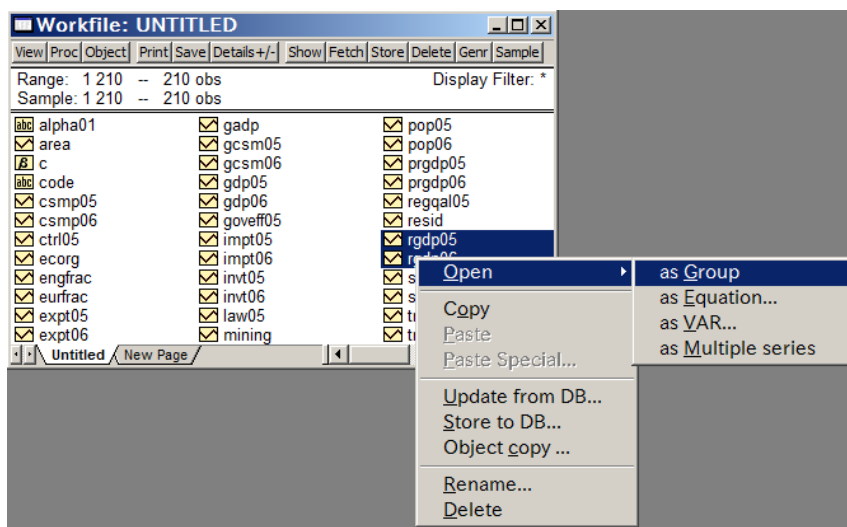
```
GENR TOKYO = PREF = 13
```

となります。論理式を AND や OR でつないで、より複雑な条件付けをすることもできます。

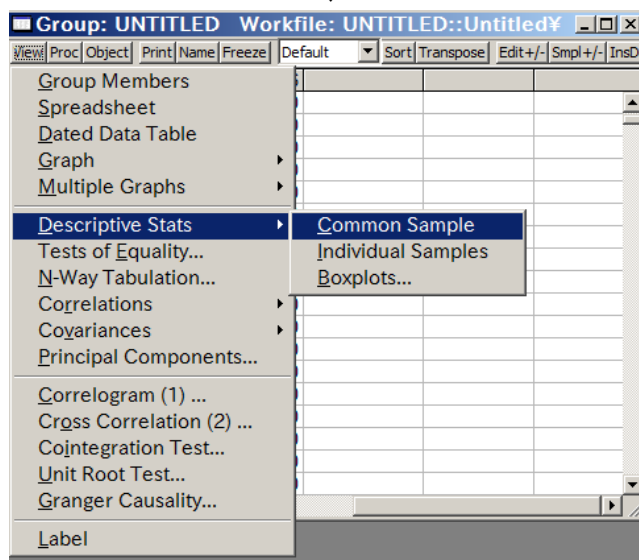
4. 標本統計量の算出

標本平均・分散、あるいは共分散や相関係数は、実証分析の手始めとして確認しておくべき統計量の代表です。

ワークファイルウィンドウで、興味のある変数をクリックし（CTRL キーを押しながら複数選択も可）、メニューバーの Show をクリック（あるいはシングルクリックで Open Group、または右クリックで Open → as Group）すると、選択した系列だけが表示されます。ここで開かれるグループウィンドウのメニューバーの View → Descriptive stats → Common Sample を選ぶと、代表的な標本統計量が表示されます。Common Sample とは、複数の変数を選んだときに、双方の変数が観測できている観測値のみを対象とした標本統計量を表示するというオプションで、Descriptive stats → Individual samples を選ぶと、それぞれの変数で得られているすべての観測値を用いて標本統計量が計算されます。欠損値の多いデータを扱う場合、注目している変数たちを選んで Common Sample の標本統計量を計算させることができます。表示させた標本統計量を選択すれば、メニューバーの Edit → Copy によって外部アプリケーションへ貼り付けできます。また、容易に想像できるように、メニューバーの View → Correlation を選ぶと相関係数行列が表示されますし、Covariance を選ぶと共分散行列が表示されます。分布を数値で表したいとき、分布の表を書きたいときにはメニューバーの View → N-way Tabulation を使います。



たとえばこうすると2つの変数が group として開かれます



こう選ぶと、欠損値を除いた標本統計量が計算されます。

ほかのを選ぶとグラフを描いたり、共分散行列を出したりします。

Group ウィンドウで、メニューバーの View → Graph からグラフを描くことができます。Line は折れ線グラフ (obs の順で)、Area はシェアのグラフ、Bar は棒グラフを表示させます。表示させたグラフが気に入らなければ、選びなおせば画面が切り替わります。散布図 (Scatter) を選ぶとさらに選択肢があります。単純な散布図 (Simple scatter)、回帰直線付きの散布図 (Scatter with regression)、ノンパラメトリック推定値付きの散布図 (Scatter with Kernel fit) などです。回帰直線付きの散布図を選ぶと、それぞれの変数を変換するかどうか決められます (対数変換、逆数など)。表示させたグラフは Object → View Options... → Save graph to disk オ

プシオンで保存できます。散布図の点を結んだようなグラフは Graph → XY line で描くことができます。ただし、時系列データの場合には自動的に横軸に時間軸が来るので、わざわざ XY グラフを使う必要はないかもしれません。Group ウィンドウで 3 個以上の変数を選んでグラフを描くときには、順番に気をつけないとへんてこりんな図が出力されます。Group ウィンドウには、CTRL キーを押しながらクリックした順番に系列が入っていき、散布図の横軸に来る変数は、一番最初にクリックされた変数 (Spreadsheet 画面で一番左に入っている変数) になります。また、1 変数しか選んでいないときには、ヒストグラムを描くことができます。各グラフをダブルクリック、または右クリックから Options を選ぶと、グラフについての設定を指定することができます。

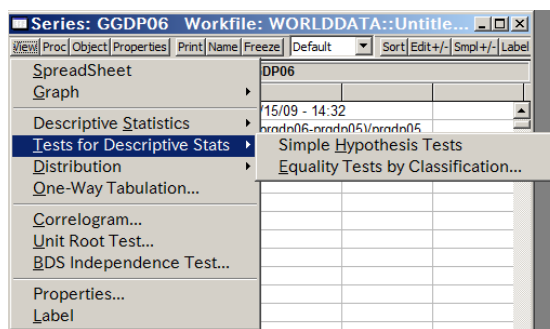
標本統計量を計算したり、グラフを描いたりすると、Group ウィンドウに表示されるものは次々と変化しますが、扱っている系列が変わっているというわけではありません。Group ウィンドウのメニューバーの view → Spreadsheet をクリックすると、Group ウィンドウを開いたときに最初に表示されるスプレッドシート画面に戻ります。つまり、メニューバーの view からさまざまなコマンドを実行することで、その系列に関する情報を次々と表示させるのが Group ウィンドウなのです。ここらへんが「オブジェクト指向」といわれる所以なのかもしれません。したがって、出力された標本統計量やグラフは表示を変えると消えてしまいます。もし他のところで使うのであれば、Copy-Paste やファイル出力などの手続きを踏まなくてはなりません。その代わり(?), このグループに名前を付けてオブジェクトとしておけば、ワークファイルの保存時にこのオブジェクトも保存されます。

同様の出力は、メニューバーの Quick → Group statistics →... からでも得ることができますが、この場合、対象となる変数名を手入力する必要があります。

もし、サンプルを区切って標本統計などを計算したい場合には、Group ウィンドウを開く前に、Workfile ウィンドウの sample から条件を入力しておきます。

5. 1 変数に関する検定

1 変数の平均や分散についての仮説検定を行うときには、Workfile ウィンドウで変数を選んで Series ウィンドウを開きます。Series ウィンドウのメニューバーの View → Tests for Descriptive Stats から検定を行います。ここで Simple hypothesis tests を選ぶと、平均値や分散についての検定を行うことができ、開いたウィンドウの Mean などに帰無仮説の値を入れると t 値と p 値が表示されます。Tests for Descriptive Stats → Equality tests by classification を選ぶと、ウィンドウに Series/Group for classify というボックスが出てきます。ここには、系列の分割に対応する系列の名前を入力します。つまり、2 グループ間の比較を行いたいのなら、どちらのグループに属するかを表す変数を Workfile 内にあらかじめ作っておく必要があります。



こう選ぶと、1 変数に関する検定を行うことができます

6. 線形回帰

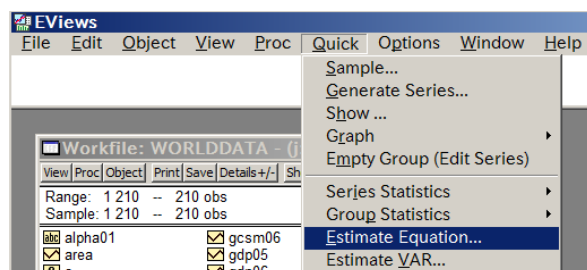
単純な重回帰分析であれば、コマンド 1 つで行うことができます。変数 Y を X_1 、 X_2 に回帰するときには、コマンドウィンドウに

```
LS Y C X1 X2
```

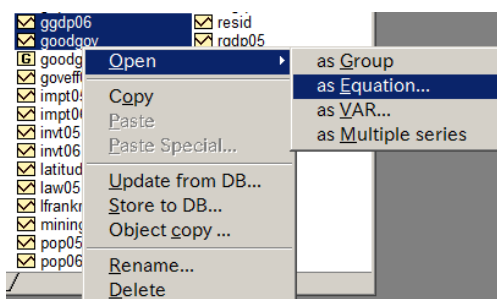
と入力してリターンキーを押せば結果が表示されます。あるいはメニューバーの Quick → Estimate Equation を選んで Equation Estimation ダイアログを開き、「Y C X1 X2」と入力して LS にチェックを入れても同様の結果を得ます。「C」は定数項を表しているから、通常の特定化では「C」は明記する必要があります。

あるいは、Workfile ウィンドウで CTRL キーを押しつつ系列を選択して白黒反転させたのち、右クリックして Open → as Equation とすると、Equation Estimation ウィンドウが、Specification タブの Equation specification ボックスに変数名が入力された形で開

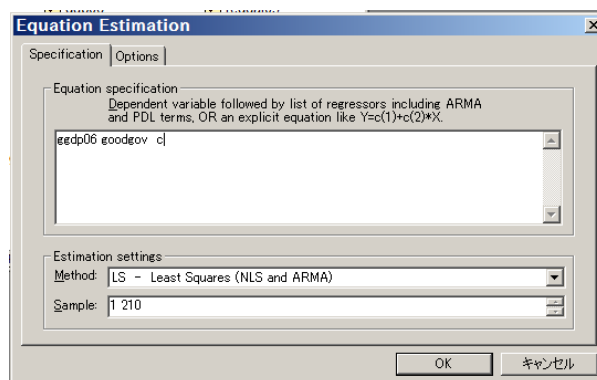
きます。このときには、CTRL キーを押しながらクリックした順番に変数名が入力され、最後に変数「c」が入ったかたちになっていますから、CTRL キーを押しつつ、被説明変数→説明変数の順序でクリックしておくこととすぐに次のステップへ移動できます。



Quick のほうからはここを選びます

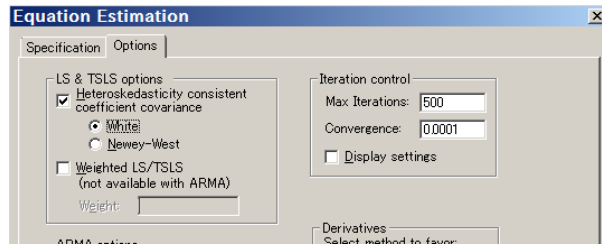


CTRL キーを押しつつクリックして選んで、右クリックでも開けます



いずれにしてもこのウィンドウを開いて、
被説明変数→説明変数の順に並べます

さて、Equation Estimation ウィンドウには option タブがあり、推定のオプションを指定できます。たとえば、heteroskedasticity にチェックを入れると heteroskedasticity-robust な標準偏差を用いた結果を表示します。オプションを指定して OK をクリックすると、推定が行われて Equation ウィンドウが開きます。Equation ウィンドウでは、行われた推定にまつわるいろいろな統計量やグラフを出力することができます。



分散不均一に頑健な標準誤差を出力させましょう

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GOODGOV	-0.002508	0.004925	-0.509219	0.6112
C	0.043577	0.003681	11.83969	0.0000

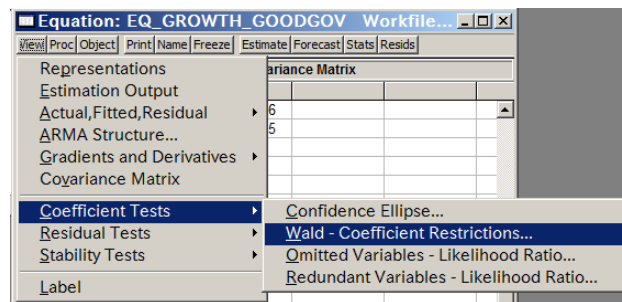
R-squared	0.002064	Mean dependent var	0.043644
Adjusted R-squared	-0.003188	S.D. dependent var	0.051874
S.E. of regression	0.051957	Akaike info criterion	-3.066455
Sum squared resid	0.512902	Schwarz criterion	-3.032523
Log likelihood	296.3797	F-statistic	0.393046
Durbin-Watson stat	1.955550	Prob(F-statistic)	0.531455

最初に推定結果の一覧表が出てきます。

開いた Equation ウィンドウのメニューバーにもさまざま出ていますが、View → Representations を選ぶと、推定式をいくつかの方法で表現してくれます。このうち、係数を c(1) などと表現しているのは係数制約についての検定を行うために有用です。View → Coefficient tests → Wald を選ぶと F 検定を行うことができます。ここで係数制約は c(1) などを用いて表現します。たとえば定数項と 1 番目の変数の係数がともにゼロという仮説は、このボックスに

$$c(1) = 0, c(2) = 0$$

と入力することによって行われます。ウィンドウには「Wald test」と書いてありますが、係数制約に関する F 検定が行われ、F 値と対応する p 値が表示されます。自分が着目している係数がどのように表現されているかは、View → Representations で確認しましょう。メニューバーの Stat をクリックすると、最初に表示された通常の回帰分析の結果が表示されるようになっているので、結果が失われたわけではありません。回帰変数を追加したいときには、Estimation をクリックすると Equation Specification ダイアログが開きますから、適当に書き直すと、新たな推定結果が表示されます。このときには以前の推定結果は消えてしまうので注意しましょう。必要があれば copy and paste で外部アプリケーションに映しておくのもよいでしょう。



係数についての仮説検定はここから

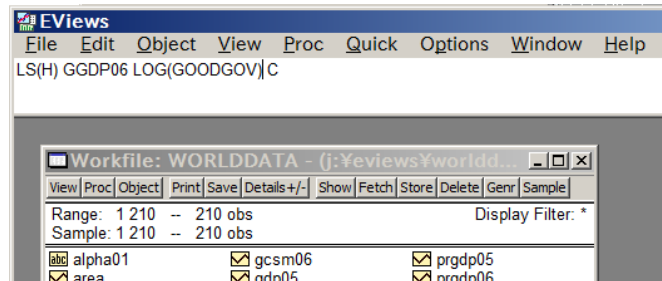
前述したように、EViews では回帰分析に使用する変数を必ずしも新たに定義する必要はありません。たとえば、変数 Y を x_1 と、 x_2 の対数値に回帰するときには

LS Y C X1 LOG (X2)

と入力すれば推定が行われますし、被説明変数を対数変換して推定するときには

LS LOG (Y) C X1 LOG (X2)

と入力すればいいことになります。もちろん、Equation Specification ダイアログから入力するときには最初の LS を入力する必要はありません。



コマンドラインから手入力してもウィンドウが開きます

LS のあとの (H) は分散不均一に頑健な標準誤差を計算させるオプション指定

推定に用いる観測値を限定するときには、Equation Specification ダイアログの Sample ボックスに論理式を記述します。たとえば変数 x_1 の値が 10 以下の観測値だけを用いて推定を行うときには「IF $x_1 \leq 10$ 」と入力します。こうすることによってサンプルが限定されますが、元のワークファイルから条件に反するものが削除されてしまったりするわけではありません。

7. 2 段階最小 2 乗推定

7.1. 推定

説明変数のいくつかが誤差項と相関をもち（内生性があり）、OLS 推定量が一致性を失うとき、操作変数法を用いて一致推定量を得ることができます。操作変数は、(1)説明変数と相関をもち、かつ、(2)誤差項と相関をもたない、という条件を満たさなければなりません。2 段階最小 2 乗推定は操作変数法の一つで、線形回帰において OLS を 2 段階行うことで一致推定量を得ることができます。

いま、内生性のある説明変数が k 個、外生性を満たす説明変数が r 個、説明変数に含まれない操作変数（除外された操作変数）が m 個ある線形回帰を考えます。被説明変数ベクトルを \mathbf{y} ($n \times 1$)、説明変数行列を \mathbf{X} ($n \times (k+r)$)、説明変数行列を \mathbf{Z} ($n \times (m+r)$) とすると、Eviews では OLS が「2 回」行われるわけではなく、係数ベクトルは

$$\beta^{\text{TSLs}} = (\mathbf{X}'[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{X})^{-1} \mathbf{X}'[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] \mathbf{y}$$

で計算されます。

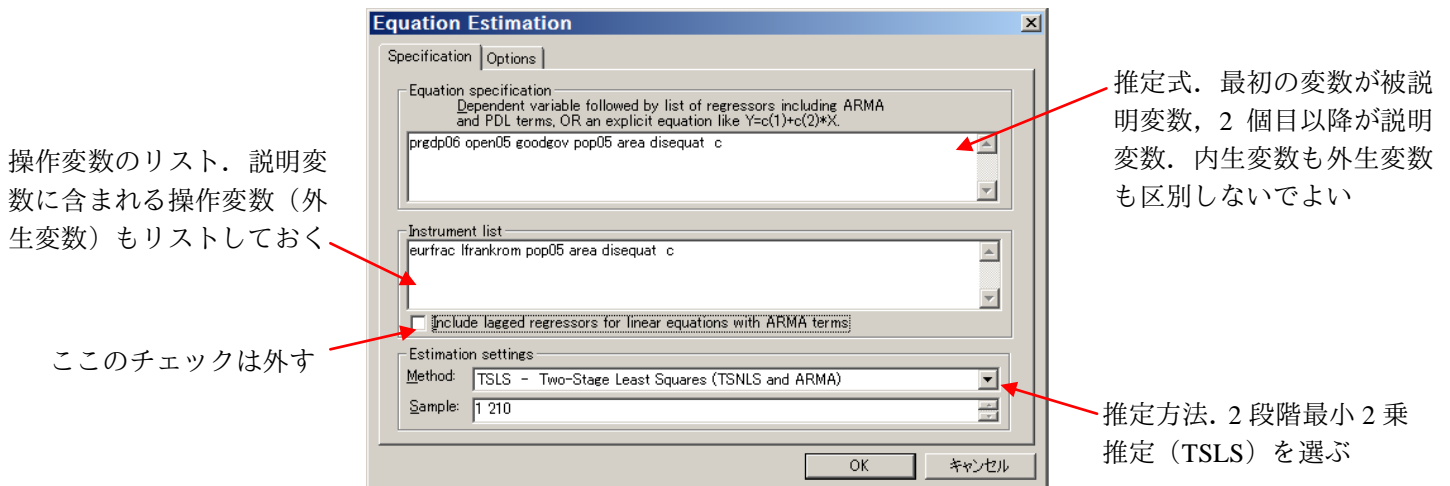
2 段階最小 2 乗推定値を得るだけなら、コマンド 1 つで行うことができます。被説明変数が Y 、内生性のある説明変数が $X1, X2$ 、外生性を満たす説明変数が $W1, W2$ 、操作変数が $Z1, Z2, Z3$ であるとき、コマンドウィンドウに

```
TSLS(h) Y X1 X2 W1 W2 C @ W1 W2 Z1 Z2 Z3
```

と入力してリターンキーを押せば結果が表示されます。あるいはメニューバーの Quick → Estimate Equation を選んで Equation Estimation ダイアログを開き、Equation Specification ボックスに「Y X1 X2 W1 W2 C」と入力して Method を TSLS とし、Instrument list ボックスに「W1 W2 Z1 Z2 Z3」と入力し、options タブを選んで左上の「Heteroskedasticity consistent coefficient covariance --- White」にチェックを入れても同様の結果を得ます。コマンドウィンドウに記述するときの「(h)」は不均一分散に頑健な分散共分散行列を計算させるオプションですから必ず書いておきましょう⁸。また、OLS の場合と同様、「C」は定数項を表していますから、通常の特定化では「C」を明記する必要があります。

OLS 推定の場合とおなじく、Workfile ウィンドウで CTRL キーを押しつつ系列を選択して白黒反転させたのち、右クリックして Open → as Equation とすると、Equation Estimation ウィンドウが、Specification タブの Equation specification ボックスに変数名が入力された形で開きます（操作変数リストは空欄のままです）。操作変数のリストを入力し、推定方法を選べば、同じ結果を得ることができます。CTRL キーを押しながらクリックした順番に変数名が入力され、最後に変数 C が入ったかたちになるので、CTRL キーを押しながら被説明変数 → 説明変数の順序でクリックすると便利なのも同じです。

⁸ 「TSLS」と「(h)」のあいだにスペースがあるとエラーが返ってきました。気をつけましょー。



2 段階最小 2 乗推定を行うためのウィンドウを開いたところ

さて、2 段階最小 2 乗推定を EViews で行うときにいくつか注意すべき点があります。第 1 は、操作変数の入力についてです。コマンドラインの「@」よりあと、Equation Estimation ダイアログの Instrument list ボックスのなかには、外生性を満たす変数をすべて記述する必要があります。計量経済パッケージによっては、内生変数を指定して、除外された操作変数を指定するという形式をとっているものもありますが、EViews では説明変数に含まれるものも含めて操作変数をすべて列挙するという形式になっています。前述の記法に従えば、 X と Z に含まれる変数を指定する形式をとっていることとなります。逆にいえば、説明変数のリストにあつて、操作変数のリストにないものが内生変数とみなされます。結果の出力にも内生変数のリストはありませんから、注意が必要です。

第 2 の点は、操作変数としての定数項の扱いです。説明変数としての定数項は明示的に「c」と入力しなければ考慮されませんが、操作変数には定数項はデフォルトで含まれます。したがって、操作変数リストに「c」を入れても入れなくても、操作変数に定数項を含んだ推定が行われます。操作変数から定数項を外すことは考えにくい、というのが理由のようです。

第 3 の点は、識別条件です。2 段階最小 2 乗推定が実行可能であるためには、除外された操作変数の数が内生変数の数を同じか、それより大きくなければなりません。もし、内生変数の数のほうが小さければ、「Order condition violated --- Insufficient instruments」というエラーメッセージが返されます。説明変数に含まれる外生変数を操作変数リストに入れ忘れていたりしたときにこのエラーメッセージが返ってくる場合があります(経験者談)。多重共線性については、前述の記法でいうところの X と Z での多重共線性が問

題になるので，説明変数間の多重共線だけでなく，操作変数間の多重共線にも注意しなければなりません。

Variable	Coefficient	Std. Error	t-Statistic	Prob.
OPEN05	3.948307	3.611342	1.093307	0.2762
GOODGOV	4.870726	1.659969	2.934228	0.0039
POP05	-2.42E-09	3.21E-09	-0.751710	0.4536
AREA	1.45E-07	5.52E-07	0.262624	0.7932
DISEQUAT	25.97117	7.633501	3.402262	0.0009
C	-3.157158	4.670100	-0.676036	0.5002

R-squared	0.633710	Mean dependent var	7.595462
Adjusted R-squared	0.619940	S.D. dependent var	11.36150
S.E. of regression	7.004251	Sum squared resid	6524.918
Durbin-Watson stat	1.951678	Second-stage SSR	8771.338

2 段階最小 2 乗推定の出力結果
読み方は OLS のときとほとんど同じです

7.2. 検定

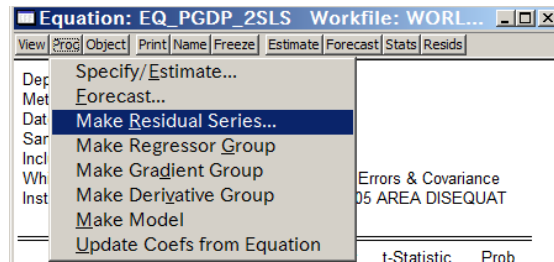
係数の制約に関する統計的検定の手続きは，OLS の場合と同じです。

内生変数が 1 つのときに操作変数の妥当性を検証する方法のひとつが，1 段階目の F 統計量を計算するというものです。おおまかにいって，除外変数の係数がすべてゼロであるという帰無仮説に対する F 統計量が 10 を超えていれば，weak instruments の問題はないと考えてよい，とされています。EViews では 1 段階目の F 統計量は自動的に出力されませんから，この F 統計量を算出するためには，内生変数を操作変数（説明変数に含まれるものも，除外されるものも含めて）に回帰した OLS 推定を行い，F 検定を実行しなければなりません。

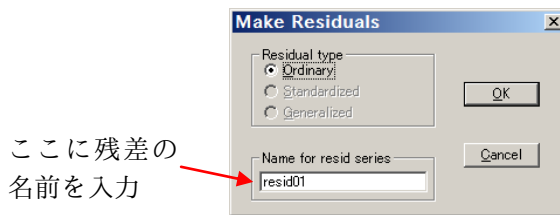
操作変数の外生性を判定する助けとなるのが，過剰識別制約検定です。この検定で用いられる統計量は Hansen の J 統計量ですが，2 段階最小 2 乗推定の場合には Sargan 統計量とも呼ばれます。EViews の 2 段階最小 2 乗推定コマンドはこの統計量を計算してくれないので，自分でコマンドを組み合わせて計算する必要があります。Sargan 統計量のアイデアは「外生変数がほんとうに外生なら，回帰の残差と外生変数は相関をもたないはずだ」というものですから，回帰の残差を外生変数に回帰して F 検定を行えばよい

こととなります⁹。したがって、次のような手続きを取ります。

1. 残差系列を作ります。2 段階最小 2 乗推定の結果が示されている Equation ウィンドウのメニュー（小さい箱のようなもの）から Proc → Make Residual Series を選びます。すると Make Residuals ウィンドウが開くので、Name for resid series ボックスに適当な名前をつけて OK とします。そうすると、Workfile ウィンドウのなかに新しく変数が追加されるはずですが。



残差系列を作る

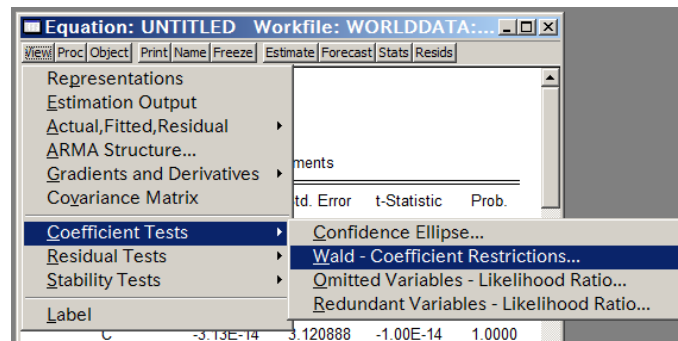


ここに残差の
名前を入力

こんなウィンドウが開くので、系列に名前をつけて保存

2. 新しく作った残差系列を被説明変数とし、2 段階最小 2 乗推定で用いた外生変数、つまり説明変数に含まれる外生変数と、除外された操作変数の両方を説明変数として、つまり Equation Estimation ダイアログの Instrument list ボックスのなかの変数すべてを説明変数として、OLS 推定を行います。このとき、分散不均一に頑健な共分散行列を用いる必要はないので、オプションは外しましょう。
3. 「除外された操作変数の係数がすべてゼロ」という帰無仮説に対する F 統計量を求めます。EViews では F 検定という名前はメニューに出てこないで、Equation ウィンドウのメニューから View → Coefficient tests → Wald を選びましょう。

⁹手続きについては、Stock and Watson [2003] Key Concept 10.6 (p.354) を参照してください。松浦・マッケンジー [2005, pp.187-189] には、Sargan 統計量の別の算出方法が載っています。



残差を操作変数に回帰して F 検定を行います

4. 表示された F 統計量に除外された操作変数の数を乗じた数がカイ 2 乗分布に従うのでそれを検定すればよい……なのですが、ここで表示された F 統計量を保存する方法がどうもなさそうなので（あったら教えてください！）、この F 統計量を（ m 倍して）メモしておいて、コマンドラインに「 $1-@cchisq(mF, m-k)$ 」と入力すると p 値が計算されるのでそれを使いましょう。 $@cchisq(x, v)$ はカイ 2 乗分布の累積分布関数で、自由度 v のカイ 2 乗分布の x における累積密度を返します。

この手続きをコマンドラインへの入力で行うと、こんな感じになります¹⁰。

```
tsls(h) y x1 x2 w1 w2 c @ w1 w2 z1 z2 z3
series iv_resid = resid
ls iv_resid z1 z2 z3 w1 w2 c
wald c(1)=0, c(2)=0, c(3)=0
1-@cchisq(mF, 1)
```

1 行目は 2 段階最小 2 乗推定，2 行目は残差系列の保存，3 行目が残差の外生変数への回帰，4 行目が除外変数についての F 検定，5 行目が J 統計量の p 値の算出です。5 行目の mF には，4 行目の検定で算出された F 統計量に操作変数の数を乗じた数を書きます¹¹。 $@cchisq(x, v)$ の 2 つ目の引数が 1 になっているのは，このケースでは除外された操作変数が 3 個，内生変数が 2 個なので，過剰識別の度数（degree of overidentification）が $3-2=1$ になっているからです。

7.3. おまけ：EViews での J 統計量

教科書に載っているくらい重要な Hansen の J 統計量（ここでは Sargan 統計量）が、

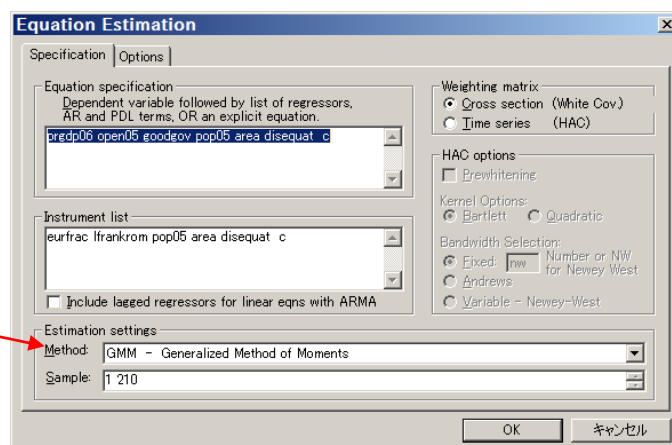
¹⁰大文字小文字はどちらでもいいです。だって DOS だもん。

¹¹ mF と書いてもエラーが返ってきます。念のため。

計量経済パッケージのくせに一発で計算されないとはどういうことじゃ、と怒るかもしれませんが（僕みたいに）が、実は EViews でも J 統計量がデフォルトで計算されることがあります。

先に述べたように、2 段階最小 2 乗推定は操作変数法の特殊ケースにあたるものですが、同時に、一般化積率法（GMM: Generalized Method of Moments）の特殊形でもあります。すなわち、線形 GMM で分散均一であるとき、効率的な GMM 推定量は 2 段階最小 2 乗推定量になるという関係にあります（Hayashi [2000], pp.225-231）。EViews にも GMM は実装されており、Equation Estimation ダイアログが開いたときに、Estimation settings ボックスで GMM を選ぶことができます。ここで適当なオプションを選ぶと 2 段階最小 2 乗推定になるか、というと、どうもそうでもないようです。さて、この GMM 推定を回してみると J 統計量はデフォルトで計算され、結果表の右下に表示されます。しかし、推定方法が違うので（efficient GMM を計算してくれるので）、前述の方法で計算した J 統計量とは別の数値が出てきます。ざんねん。

ここで GMM を選んでいるので、右上のオプションも変化している




Equation: EQ_PGDP_2SLS Workfile: WORL...

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: PRGDP06
 Method: Generalized Method of Moments
 Date: 06/30/09 Time: 16:56
 Sample (adjusted): 3 210
 Included observations: 139 after adjustments
 White Covariance
 Simultaneous weighting matrix & coefficient iteration
 Convergence achieved after: 1 weight matrix, 2 total coef iterations
 Instrument list: EURFRAC LFRANKROM POP05 AREA DISEQUAT
 C

Variable	Coefficient	Std. Error	t-Statistic	Prob.
OPEN05	3.948307	3.532540	1.117696	0.2657
GOODGOV	4.870726	1.623747	2.999684	0.0032
POP05	-2.42E-09	3.14E-09	-0.768479	0.4436
AREA	1.45E-07	5.40E-07	0.268483	0.7887
DISEQUAT	25.97117	7.466931	3.478159	0.0007
C	-3.157158	4.568195	-0.691117	0.4907

R-squared	0.633710	Mean dependent var	7.595462
Adjusted R-squared	0.619940	S.D. dependent var	11.36150
S.E. of regression	7.004251	Sum squared resid	6524.918
Durbin-Watson stat	1.951678	J-statistic	1.72E-29

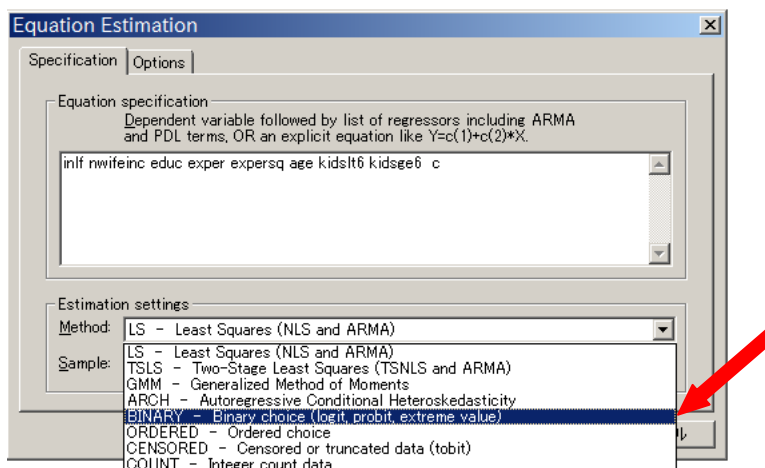


右下に J-statistics があります

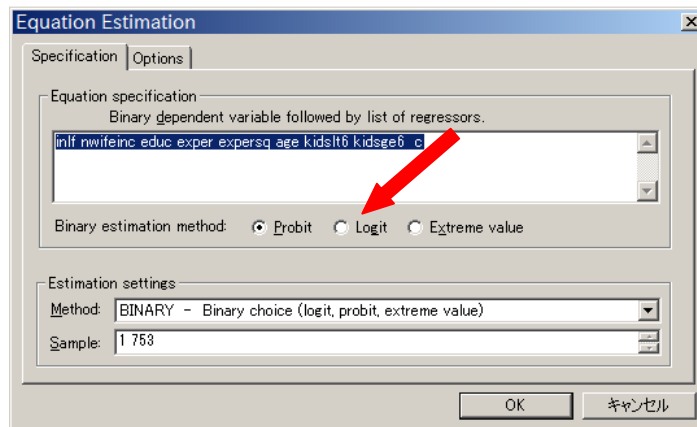
8. 離散変数モデル

被説明変数が離散的な値しかとらないとき、典型的には2通りの値しかとらないようなモデルを広く離散変数（離散選択）モデルと呼びます。被説明変数が離散的であろうとも OLS 推定を行うこと自体はできるのですが、被説明変数が Yes-No のような質的な変数のときはとくに、係数の解釈がややこしくなること、当てはめ値が任意の実数を取りうるのもっともらしくないこと、から、離散変数モデルを使うことがあります。被説明変数が2値変数のときには、Probit モデルと logit モデルがよく使われます。この2つのモデルは、probit モデルが誤差項に正規分布を仮定しているのに対し、logit モデルがロジスティック分布を仮定しているという違いはあるものの、推定結果に大きな差はありません（推定される係数の値はかなり異なりますが）。Probit と logit の違いは、多項選択モデルへの拡張のしやすさにあると思ったほうがよいのではないのでしょうか。

Eviews での推定方法は OLS とよく似ています。Workfile ウィンドウで CTRL キーを押しつつ系列を選択して白黒反転させたのち、右クリックして Open → as Equation とすると、Equation Estimation ウィンドウが、Specification タブの Equation specification ボックスに変数名が入力された形で開きます。CTRL キーを押しながらクリックした順番に変数名が入力され、最後に変数 c が入ったかたちになるので、CTRL キーを押しながら被説明変数→説明変数の順序でクリックすると便利なのも同じです。2値変数モデルのばあい、ここで Estimation settings から BINARY を選びます。



すると、Equation specification ボックスのところにオプションが出てくるので、probit, logit, extreme からひとつ選びましょう。Options では頑健な共分散行列を選んでおくとよいでしょう。



出力結果も OLS とよく似ています。

Equation: UNTITLED Workfile: MROZ::Untitled¥

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: INLF
 Method: ML - Binary Probit (Quadratic hill climbing)
 Date: 10/19/10 Time: 14:38
 Sample: 1 753
 Included observations: 753
 Convergence achieved after 4 iterations
 QML (Huber/White) standard errors & covariance

Variable	Coefficient	Std. Error	z-Statistic	Prob.
NWIFEINC	-0.012024	0.005307	-2.265619	0.0235
EDUC	0.130905	0.025802	5.073420	0.0000
EXPER	0.123348	0.018841	6.546702	0.0000
EXPER SQ	-0.001887	0.000600	-3.143466	0.0017
AGE	-0.052853	0.008348	-6.331457	0.0000
KIDSLT6	-0.868329	0.116126	-7.477438	0.0000
KIDSGE6	0.036005	0.045266	0.795414	0.4264
C	0.270077	0.504839	0.534976	0.5927

Mean dependent var	0.568393	S.D. dependent var	0.495630
S.E. of regression	0.425945	Akaike info criterion	1.087124
Sum squared resid	135.1646	Schwarz criterion	1.136251
Log likelihood	-401.3022	Hannan-Quinn criter.	1.106050
Restr. log likelihood	-514.8732	Avg. log likelihood	-0.532938
LR statistic (7 df)	227.1420	McFadden R-squared	0.220581
Probability(LR stat)	0.000000		

Obs with Dep=0	325	Total obs	753
Obs with Dep=1	428		

9. 打ち切りモデル

被説明変数がある限られた範囲の値しか取らない状況、あるいは、なんらかの条件に当てはまったとき（当てはまらなかったとき）にはデータが観測できない状況があります。このような被説明変数を広く、制限従属変数（limited dependent variable）といい、なんらかの形で被説明変数が観測できるときには打ち切られた（censored）、被説明変数が観測できないときには切断された（truncated）といいます。データが打ち切られる、あるいは切断される観測できるための条件はあらかじめ分かっていることもありますし、分かっていることもあります。また、その条件が個々の観測値によって異なるばあいも、定数のばあいも、確率的に決まるばあいも考えられます。このような状況を扱う（同時体系で表される）回帰モデルを、広い意味で Tobit モデルと呼びます。

Tobit モデルでも、離散変数モデルと同じく潜在変数（latent variable）を想定し、潜在変数が説明変数の線形結合（1 次関数）で表現できると考えたうえで、潜在変数が一定の条件を満たすときにはそのまま観察されるが、そうでないときには定数などとして観測されると考えます。

打ち切りデータの一例は労働時間です。（家庭外での）労働時間は通常はマイナスの値をとらないと考えられます¹²から、労働時間はゼロという閾値によって左に打ち切られた（left censored）となります。いま、有配偶女性の労働供給時間を被説明変数とする回帰分析を考えましょう（Wooldridge 2002, pp.527-528, Example 16.3）。OLS と同じく、Equation Estimation ウィンドウに、Specification タブの Equation specification ボックスに変数名を入力した形で開きます。

左で打ち切られているから、Left のほうにだけ、閾値の 0 を入力する

閾値は分かっているので、actual censoring を選ぶ

Censored を選ぶ

¹² 1 日の労働時間であれば 24 時間以上の値はとりませんが、こちらは通常は無視します。

Estimation settings の Method で Censored を選ぶと、そのうえにオプションが出てきます。労働時間のばあいは被説明変数が閾値 0 で左で打ち切られているのですから、Left のほうに閾値の 0 を入力します。右の閾値は空欄にしておきます。また、閾値が既知なので、entered as のところは Actual censoring value のほうを選びます。その上に、誤差項の分布形のオプションがありますが、通常は正規分布を仮定するので、デフォルトで選ばれている Normal のままでよいでしょう。

OK とすると、推定結果が出力されます。推定された係数は、潜在変数を規定する線形結合の係数ですから、解釈には注意しましょう。

Equation: UNTITLED Workfile: MROZ::Untitled*				
View Proc Object Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: HOURS				
Method: ML - Censored Normal (TOBIT) (Quadratic hill climbing)				
Date: 11/09/10 Time: 11:12				
Sample: 1 753				
Included observations: 753				
Left censoring (value) at zero				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
	Coefficient	Std. Error	z-Statistic	Prob.
NWIFEINC	-8.814243	4.459099	-1.976687	0.0481
EDUC	80.64561	21.58323	3.736493	0.0002
EXPER	131.5643	17.27939	7.613943	0.0000
EXPEXSQ	-1.864158	0.537662	-3.467156	0.0005
AGE	-54.40501	7.418501	-7.333693	0.0000
KIDSLT6	-894.0218	111.8780	-7.991040	0.0000
KIDSGE6	-16.21800	38.64139	-0.419705	0.6747
C	965.3053	446.4361	2.162247	0.0306
Error Distribution				
SCALE:C(9)	1122.022	41.57910	26.98524	0.0000
R-squared	0.273365	Mean dependent var	740.5764	
Adjusted R-squared	0.265552	S.D. dependent var	871.3142	
S.E. of regression	746.7157	Akaike info criterion	10.16758	
Sum squared resid	4.15E+08	Schwarz criterion	10.22285	
Log likelihood	-3819.095	Hannan-Quinn criter.	10.18887	
Avg. log likelihood	-5.071839			
Left censored obs	325	Right censored obs	0	
Uncensored obs	428	Total obs	753	

10. 標本選択モデル：Heckit

一定の条件を満たさない観測値（データの一部）は観測されず，被説明変数の値の情報を利用できない．つまり母集団からランダムサンプリングした標本のうちの一部分しか利用できない状況も考えられます．このようなとき，標本選択（sample selection）が起きていると言います．ここで扱う標本選択モデルでは，説明変数が観測できていることに注意しましょう．

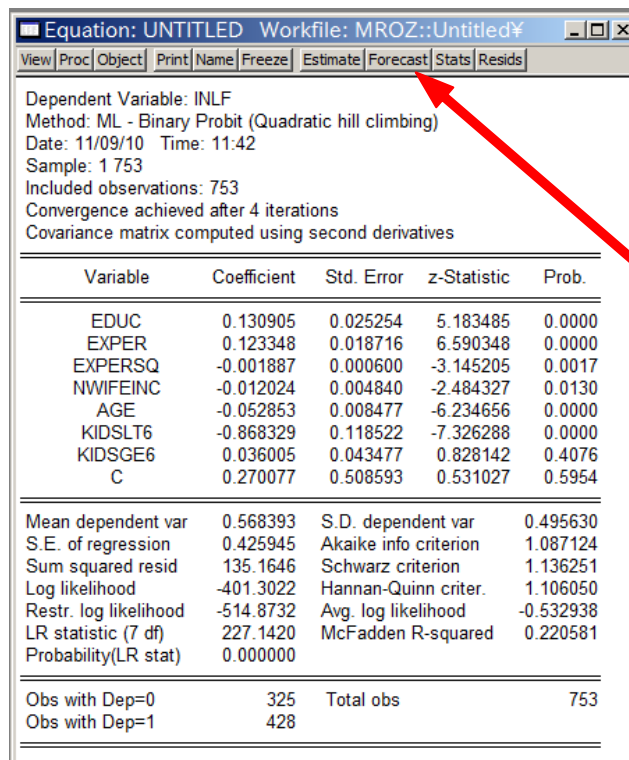
標本選択の典型例は，企業が申し出る（offer）賃金率の推定です．労働者は企業が offer する賃金率を受け入れて働くか，受け入れずに働かないか（take it or leave it）どちらかであるとすれば，観測される賃金率は受け入れられた賃金率だけですから，受け入れられなかった賃金率は観測されません．

標本選択モデルの推定方法として有名なのは，Heckman の 2 段階推定と呼ばれる推定方法です．Tobit にならって(?), Heckit と呼ばれることもあります．Heckman の 2 段階推定では，(1) 全ての観測値を使って，被説明変数が観測できるかできないかの probit 推定を行い，逆ミルズ比の推定値を計算し，(2) 被説明変数が観測される観測値を使って，説明変数に逆ミルズ比の推定値を追加した最小 2 乗推定を行う，という手続きを踏みます．1 段階目では，被説明変数が観測されない観測値についてもなんらかのデータが観測できていることに注意しましょう．一般に，2 段階目の OLS で用いられる説明変数は，1 段階目の Probit で用いられる説明変数の部分集合になっており，2 段階目の OLS で用いられない変数が観測の有無を識別するのに用いられています．

賃金率の推定を考えましょう．企業が申し出る賃金率は，労働者の教育水準と経験によって決まっているとします．他方，労働者が賃金を受け入れるかどうか（留保賃金率の水準）は，非労働所得・年齢・子どもの数にも依存すると仮定します．このとき，Heckman の 2 段階推定では，まず 1 段階目で，賃金率が観測されるかどうか（働いているかどうか）の Probit を，教育水準・経験，非労働所得・年齢・子どもの数に回帰し，逆ミルズ比（正規分布の密度関数を分布関数で除したもの）を推計します．2 段階目では，賃金率を被説明変数とし，教育水準・経験と，推計された逆ミルズ比を説明変数とする OLS を行います．ここで，1 段階目で用いられた非労働所得・年齢・子どもの数は，観測されるかどうかを識別するのに用いられており，賃金を観測できていないひとたちについてもデータが利用可能である必要があります．もし，働いている人たちのデータしかなく，働いていない人のデータがなければ，ここでの標本選択モデルを利用することはできません．また，2 段階目の OLS に含まれず，かつ，1 段階目の Probit に利用できるような変数がすべての人について観測できていないときにも，標本選択モデルは利用できません．つまり，企業の申し出る賃金率が，非労働所得・年齢・子どもの数にも依存すると考えられるときには，ここにあるデータだけでは Heckit は行えません．

さて、Eviews 5.1. では Heckman の 2 段階推定は実装されていませんから、Probit と OLS をべつべつに行う必要があります。なお、1 段階目と 2 段階目をいっぺんに最尤推定する方法もあり、標本選択モデルが備えられている Stata では、「2 段階」推定するかどうかをオプション指定する必要があります。

さて、まず、Probit 推定すると、このような出力を得ます。



Equation: UNTITLED Workfile: MROZ::Untitled¥

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: INLF
Method: ML - Binary Probit (Quadratic hill climbing)
Date: 11/09/10 Time: 11:42
Sample: 1 753
Included observations: 753
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

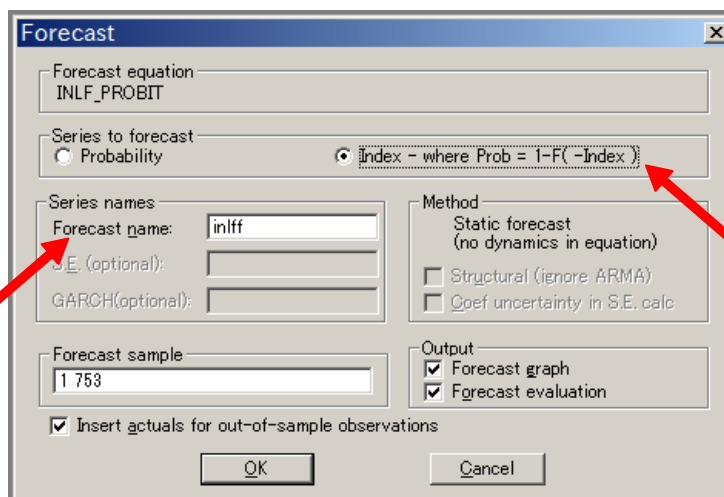
Variable	Coefficient	Std. Error	z-Statistic	Prob.
EDUC	0.130905	0.025254	5.183485	0.0000
EXPER	0.123348	0.018716	6.590348	0.0000
EXPERSQ	-0.001887	0.000600	-3.145205	0.0017
NWIFEINC	-0.012024	0.004840	-2.484327	0.0130
AGE	-0.052853	0.008477	-6.234656	0.0000
KIDSLT6	-0.868329	0.118522	-7.326288	0.0000
KIDSGE6	0.036005	0.043477	0.828142	0.4076
C	0.270077	0.508593	0.531027	0.5954

Mean dependent var	0.568393	S.D. dependent var	0.495630
S.E. of regression	0.425945	Akaike info criterion	1.087124
Sum squared resid	135.1646	Schwarz criterion	1.136251
Log likelihood	-401.3022	Hannan-Quinn criter.	1.106050
Restr. log likelihood	-514.8732	Avg. log likelihood	-0.532938
LR statistic (7 df)	227.1420	McFadden R-squared	0.220581
Probability(LR stat)	0.000000		

Obs with Dep=0	325	Total obs	753
Obs with Dep=1	428		

Forecast をクリック

逆ミルズ比は当てはめ値みたいなものなので、Equation ウィンドウの Forecast をクリックします。Probit なので、確率の当てはめ値も選べますが、ここでは $x\beta$ を選びます。



Forecast

Forecast equation
INLF_PROBIT

Series to forecast
 Probability
 Index - where Prob = 1-F(-Index)

Series names
Forecast name: inlff
S.E. (optional):
GARCH(optional):

Method
Static forecast (no dynamics in equation)
 Structural (ignore ARMA)
 Coef uncertainty in S.E. calc

Forecast sample
1 753

Output
 Forecast graph
 Forecast evaluation

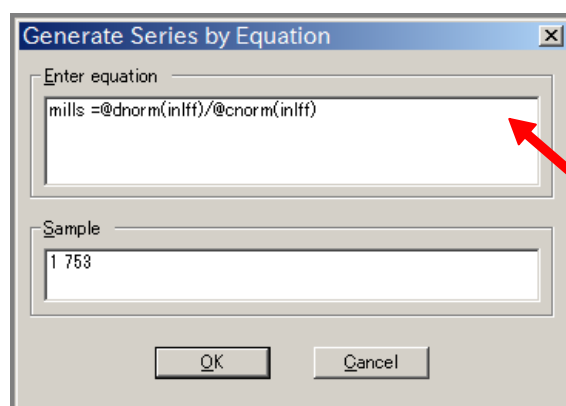
Insert actuals for out-of-sample observations

OK Cancel

Index を選ぶ

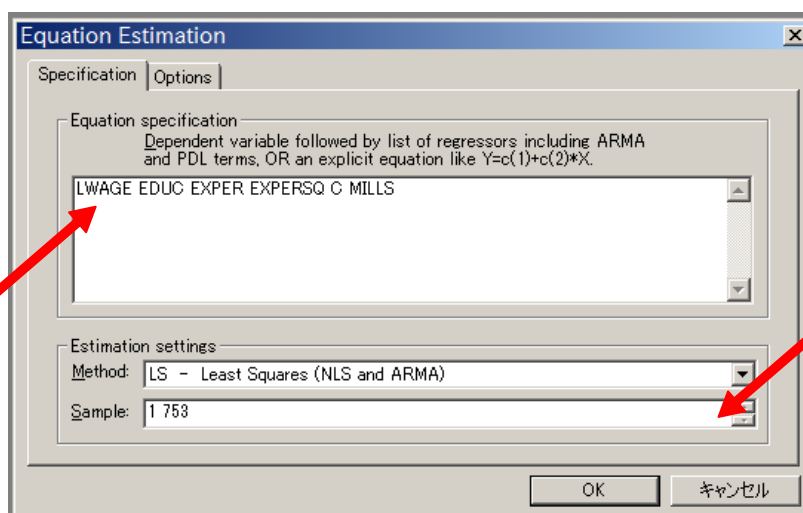
当てはめ値が入る系列名を指定する。デフォルトのままでもよい。

ここでは、 $x\beta$ の当てはめ値に inlff という名前をつけています。次に、メニューバーの Quick→Generate Series... を開いて、逆ミルズ比を新たな変数として作成します。もちろん、メニューバーのすぐ下にあるコマンドウィンドウにコマンドラインを入力してもかまいません。逆ミルズ比は正規分布の密度関数と分布関数の比であり、Eviews では正規分布の密度関数は@dnorm(), 分布関数は@cnorm() で表されますから、この関数の引数に inlff を指定すれば逆ミルズ比の推計系列を作成することができます。式で書けば、mills = @dnorm(inlff)/@cnorm(inlff) となります。



ここに式を入力

こうして作成した逆ミルズ比を説明変数に加えて、通常の OLS 推定を行います。ここでは mills という名前をつけているので、これを説明変数に加えます。推定の対象となるのは賃金率が観測された観測値だけですが、観測されていない観測値がデータ上 N.A.となっていれば、自動的に推定からは落とされるので、sample は全体を指定しておいてかまいません。



説明変数に mills を追加している

働いていない人の賃金率データが N.A. になっていれば自動的に推定から落とされるので、そのままよい

推定自体はただの OLS です。

Equation: LWAGE_HECKIT Workfile: MROZ::Unti...									
View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: LWAGE									
Method: Least Squares									
Date: 11/09/10 Time: 12:00									
Sample (adjusted): 1 428									
Included observations: 428 after adjustments									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
EDUC	0.109066	0.015610	6.987080	0.0000					
EXPER	0.043887	0.016353	2.683688	0.0076					
EXPER SQ	-0.000859	0.000441	-1.946356	0.0523					
C	-0.578103	0.306723	-1.884773	0.0601					
MILLS	0.032262	0.134388	0.240066	0.8104					
R-squared	0.156935	Mean dependent var	1.190173						
Adjusted R-squared	0.148963	S.D. dependent var	0.723198						
S.E. of regression	0.667162	Akaike info criterion	2.040046						
Sum squared resid	188.2795	Schwarz criterion	2.087466						
Log likelihood	-431.5698	F-statistic	19.68521						
Durbin-Watson stat	1.958436	Prob(F-statistic)	0.000000						

ちなみに、逆ミルズ比を追加せずに推定するとこのような結果になります。

Equation: UNTITLED Workfile: MROZ::Untitled¥									
View	Proc	Object	Print	Name	Freeze	Estimate	Forecast	Stats	Resids
Dependent Variable: LWAGE									
Method: Least Squares									
Date: 11/09/10 Time: 11:18									
Sample (adjusted): 1 428									
Included observations: 428 after adjustments									
Variable	Coefficient	Std. Error	t-Statistic	Prob.					
EDUC	0.107490	0.014146	7.598333	0.0000					
EXPER	0.041567	0.013175	3.154906	0.0017					
EXPER SQ	-0.000811	0.000393	-2.062833	0.0397					
C	-0.522041	0.198632	-2.628179	0.0089					
R-squared	0.156820	Mean dependent var	1.190173						
Adjusted R-squared	0.150855	S.D. dependent var	0.723198						
S.E. of regression	0.666420	Akaike info criterion	2.035509						
Sum squared resid	188.3052	Schwarz criterion	2.073445						
Log likelihood	-431.5990	F-statistic	26.28616						
Durbin-Watson stat	1.960988	Prob(F-statistic)	0.000000						

このケースでは、2段階目の OLS で逆ミルズ比の係数はゼロと統計的に有意に異ならな
いので、標本選択は大きな問題となっていないことになります。じっさい、ただの OLS
と比較しても、そのほかの係数推定値に大きな差はありません。