

1 章 : はじめての EViews

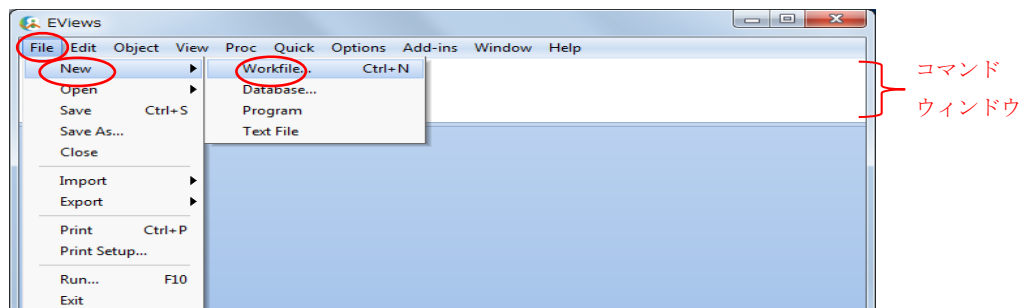
ここでは分析を行うにあたって、代表的なツールの 1 つとして EViews について解説しよう。EViews は、時系列分析に強みを持つ統計ソフトであり、その使い易さ、また高度な分析に対応できることから、官公庁を中心に広く用いられている。

1. データの入力と保存

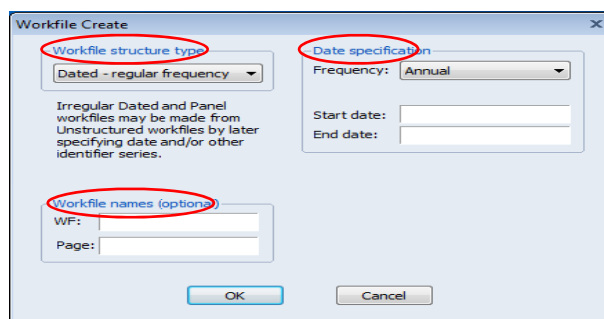
EViews では、データを特有のファイル形式である **Workfile** に変換してから操作を行う。ここでは **Workfile** を新規作成し、エクセルからデータを読み込んだうえで、保存する方法を説明する。使用するデータは、1947 年から 2012 年までの米国の四半期データとする。データファイル RGDP.xls は、HP からダウンロードできる。

1.1. Workfile の新規作成

EViews を起動させると、下図のウィンドウが出てくるので、左上の「File」→「New」→「Workfile」を選択する。



そうすると、Workfile Create ウィンドウが表示されるので、ここでデータの種類を指定する。下図をみると、左上に Workfile structure type、右上に Date specification、左下に Workfile names (Workfile の名前を設定するためのボックス) とある。



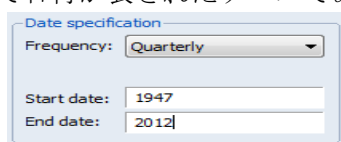
¹ 本稿は『実証分析のための時系列解析』（著者：エンダース、翻訳者：新谷元嗣、藪友良）の付属資料として、作成しました。データは全て付属 Website (<http://www.fbc.keio.ac.jp/~tyabu/enders>) から入手できます。コメントは tyabu@fbc.keio.ac.jp まで送って頂けると幸いです。

まず、Workfile structure type ではデータの種類を指定する。横断面（クロスセクション）であれば Unstructured/Undated、時系列であれば Dated-regular frequency、パネルであれば Balanced Panel を選ぶ。デフォルト設定は Dated-regular frequency である。ここでは、時系列データを分析するため、変更しないでそのままにしておこう²。

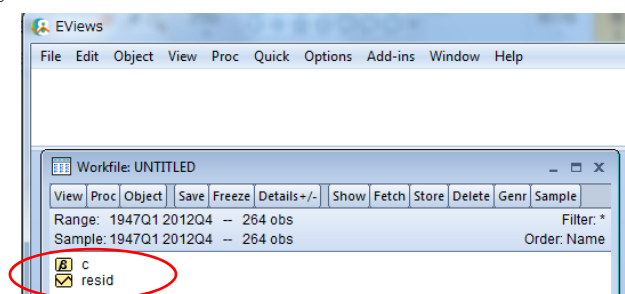
次に、Date specification で、観察頻度（frequency）、分析期間を指定する。下表には、観察頻度の種類をまとめている。データは年次(Annual)、四半期（Quarterly）、月次(Monthly) 週次(Weekly)がほとんどである。もし日次(Daily)で土日が欠損値となっていれば Daily[5 day weeks]、土日も含むなら Daily[7 day weeks]を選べばよい。

	観察頻度
Annual	年次
Semi-annual	半年次
Quarterly	四半期
Monthly	月次
Weekly	週次
Daily[5 day weeks]	日次(土日を除く)
Daily[7 day weeks]	日次
Integer date	整数日付

期間は始期（Start date）と終期（End date）をボックスに入力して特定する。ここでは、1947年から2012年までの四半期データ RGDP.xls を分析するため、Frequency は Quarterly、Start date は 1947、End date は 2012 とする。整数日付(Integer date)は、1（期）、2（期）、...、100（期）のように整数によって日付が表されたケースである。



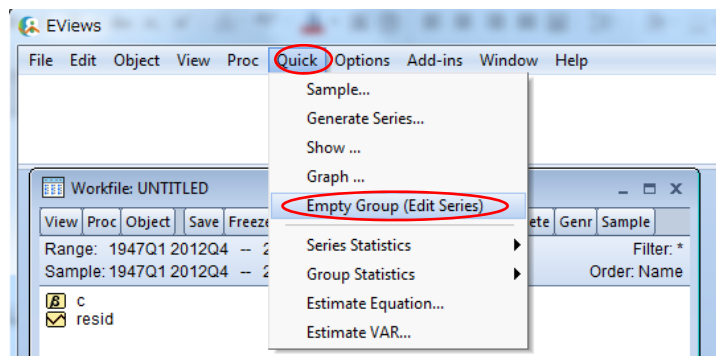
最後に、OK を押すと、ダイアログが閉じて、c と resid のみが表示された新規の Workfile が作成される。画面は下図のようになるが、変数名欄において、c は定数、resid は残差を意味する。残差はモデルを推定した後に、データが入力されるため、この時点で情報は何も入っていない。



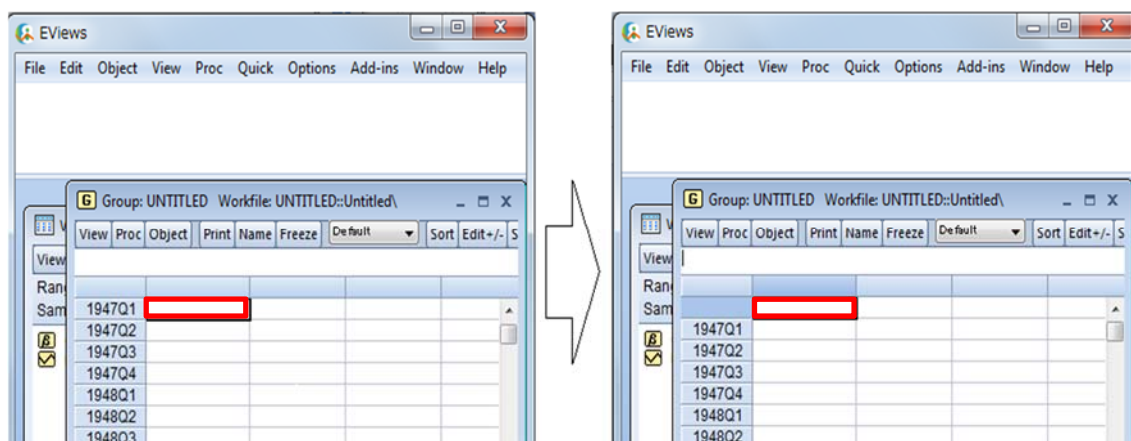
² なお、クロスセクションデータを選択した場合は同ウィンドウ右の Data range ボックスにてサンプルサイズ (Observations) を、パネルデータの場合は Panel specification で頻度、期間、被説明変数の数を指定する。

1.2. Excel から EViews へのデータ入力

新規の Workfile を作成したら、RGDP.xls からデータを入力しよう。まず EViews のメインメニューから「Quick」→「Empty Group (Edit Series)」を選択する（下図参照）。



すると、Workfile のスプレッドシートウィンドウが表示される（左下図）。ここで注意しておきたいのは、最上段にはデータラベル（データ名）の入力欄があるが見えなくなっている点である。このため、（スプレッドシートをクリックしてから）カーソルを上に戻かし、最上段を表示したうえで、ラベルを含めたデータをエクセルからコピーして貼り付ける（右下図）³。



ホームページのデータ RGDP.xls ファイルを開いてみよう。そして、ラベルを含めてデータをコピーする（下図参照）。ただし、エクセルデータ 1 列目の時点情報はコピーしないでよい⁴。このデータでは、rgdp は実質 GDP、gdp は名目 GDP、rcons は実質消費額、rgovt は実質政府支出、rinv は実質投資額、potential は潜在実質 GDP を表している。Potential は、1947、1948 年は空白になっているが、これはデータがないことを意味している。

³ EViews は英語のアプリケーションであるため、ラベルには半角英数字を使用する必要がある。また、EViews では大文字と小文字の区別がされないことにも留意したい。

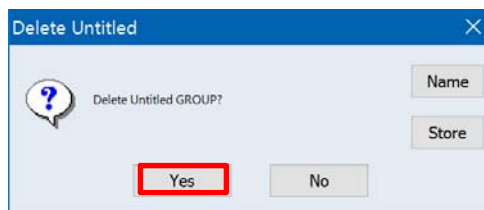
⁴ エクセルの時点情報は、1947-01-01 からスタートしているが、これは 1947 年の第一四半期を意味している。同様に、1947-04-01 は 1947 年第二四半期である。

	A	B	C	D	E	F	G	H	I	J
1		rgdp	gdp	rcons	rgovt	rinv	potential			
2	1947-01-01	1770.7	237.2	1130.9	407.250	188.675				
3	1947-04-01	1768	240.4	1149.7	409.821	173.485				
4	1947-07-01	1766.5	244.5	1153.4	413.577	167.924				
5	1947-10-01	1793.3	254.3	1153.8	407.464	202.712				
6	1948-01-01	1821.8	260.3	1159.6	415.526	224.682				
7	1948-04-01	1855.3	267.3	1173.0	435.526	239.489				
8	1948-07-01	1865.3	273.8	1174.8	444.004	242.053				
9	1948-10-01	1868.2	275.1	1184.1	462.670	230.008				
10	1949-01-01	1842.2	269.9	1186.0	475.894	194.233	1867			
11	1949-04-01	1835.5	266.2	1204.1	497.437	167.057	1885.3			
12	1949-07-01	1856.1	267.6	1206.9	502.434	181.690	1903.9			
13	1949-10-01	1838.7	265.2	1224.5	491.757	171.768	1922.9			
14	1950-01-01	1913	275.2	1244.9	483.529	214.858	1942.5			
15	1950-04-01	1971.2	284.5	1265.4	491.441	238.901	1962.9			
16	1950-07-01	2048.4	301.9	1330.3	482.007	260.418	1984.1			

そして、これを EViews のスプレッドシートウィンドウに貼り付ける。Workfile ウィンドウを見ると、変数名がリスト化されて表示されている（下図参照）。potential に関しては、1947Q1~1948Q4 は NA になっている。これは Not Available の略であり、データが利用できないことを意味している。

	RGDP	GDP	RCONS	RGOVT	RINV	POTENTIAL
1947Q1	1770.700	237.2000	1130.900	407.2500	188.6750	NA
1947Q2	1768.000	240.4000	1149.700	409.8210	173.4850	NA
1947Q3	1766.500	244.5000	1153.400	413.5770	167.9240	NA
1947Q4	1793.300	254.3000	1153.800	407.4640	202.7120	NA
1948Q1	1821.800	260.3000	1159.600	415.5260	224.6820	NA
1948Q2	1855.300	267.3000	1173.000	435.5260	239.4890	NA
1948Q3	1865.300	273.8000	1174.800	444.0040	242.0530	NA
1948Q4	1868.200	275.1000	1184.100	462.6700	230.0080	NA
1949Q1	1842.200	269.9000	1186.000	475.8940	194.2330	1867.000
1949Q2	1835.500	266.2000	1204.100	497.4370	167.0570	1885.300
1949Q3	1856.100	267.6000	1206.900	502.4340	181.6900	1903.900
1949Q4	1838.700	265.2000	1224.500	491.7570	171.7680	1922.900
1950Q1	1913.000	275.2000	1244.900	483.5290	214.8580	1942.500
1950Q2	1971.200	284.5000	1265.400	491.4410	238.9010	1962.900

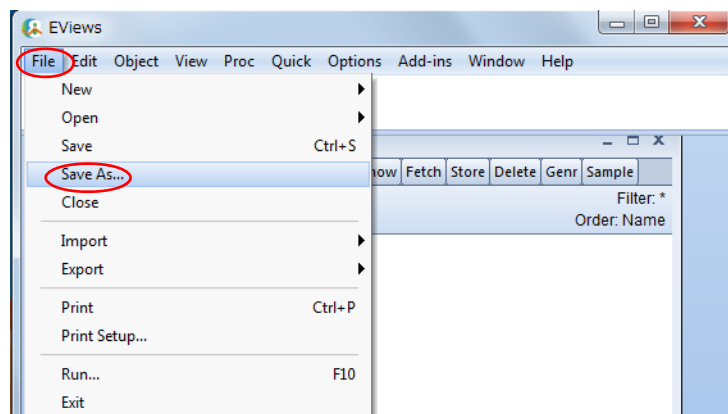
データが読み込まれたことを確認したら、スプレッドシートウィンドウは閉じてよい。スプレッドシートを閉じても、Workfile ウィンドウに保存されているため、データが失われることはない（ここで「Delete Untitled Group?」とダイアログが表示されるが、Yes として問題ない）。



1.3 データの保存

EViews でデータを入力することができたら、Workfile 形式でデータを保存してみよう。メインメニューの「File」→「Save as」を選ぶと、名前を付けて保存することができる（下

図参照)。ウィンドウが出てきたら、保存する場所を選び、ファイル名を入力する。ファイルの種類はデフォルトの EViews Workfile(*.wfl)とする。また上書き保存をする場合には、「File」→「Save」を選択する。



このように Workfile 形式でデータを保存しておくことで、このファイルを開けば、既にデータが読み込まれた状況になっており、面倒なデータ入力操作を必要はない。

2. データの操作

2.1 変数の変換

分析を行う際には、変数を加工して扱う場合がある。EViews では、Workfile に保存されている系列のデータを変換し、新たな系列を作成できる。新たな変数を定義するには、コマンドウィンドウに

`genr 新しい変数の名前 = 定義式`

と入力し、Enter を押して実行すればよい (genr は generate の略)。たとえば、変数 X の 2 乗を、X2 と名付けて定義するなら

`genr X2 = X*X`

と入力する。また、X の 2 乗は

`Genr X2=X^2`

としてもよい。同様に、X の 3 乗は X^3 とする。X の自然対数を、LX と名付けて定義するなら、

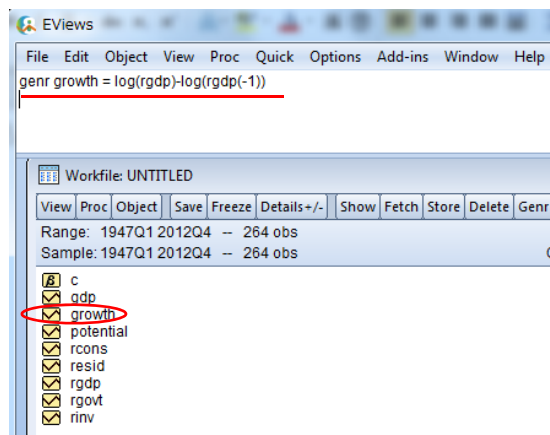
`genr LX = log(X)`

と入力すればよい。

先ほど、読み込んだデータを使って、データ変換の手順を確認しよう。たとえば、実質 GDP の対数の差 ($\log(\text{rgdp}_t) - \log(\text{rgdp}_{t-1})$) を growth という名前で定義したいとする。前期の実質 GDP は $\text{rgdp}(-1)$ として表されるため、コマンドは

`genr growth = log(rgdp)-log(rgdp(-1))`

とすればよい。また、階差(1期前との差)をとる関数 $d(X)$ を用いて、 $\text{genr growth} = d(\log(\text{rgdp}))$ と書くこともできる。コマンドを入力すると、workfile ウィンドウに、新たな変数 **growth** が表示される。



このように **genr** コマンドを、コマンドウィンドウに入力、**Enter** を押すと、Workfile に新たな変数が定義される。下表では、変数を定義する場合に便利な関数形をまとめているので参考にしてもらいたい。

関数	説明
$\log(X)$	自然対数
$\exp(X)$	指数
$\text{abs}(X)$	絶対値
$\text{sqr}(X)$	平方根
$d(X)$	階差 (前期差) $X_t - X_{t-1}$
$\text{@pch}(X)$	前期比 $(X_t - X_{t-1}) / X_{t-1}$
$\text{@pchy}(X)$	前年同期比 四半期データなら $(X_t - X_{t-4}) / X_{t-4}$

ダミー変数を作るには、新しい変数の名前と論理式を等式で結ぶ。たとえば、変数 **growth** が 0 を下回った場合に 1、その他を 0 とするダミー変数を **D1** として定義するなら、

$$\text{genr D1} = \text{growth} < 0$$

とコマンドを入力する。

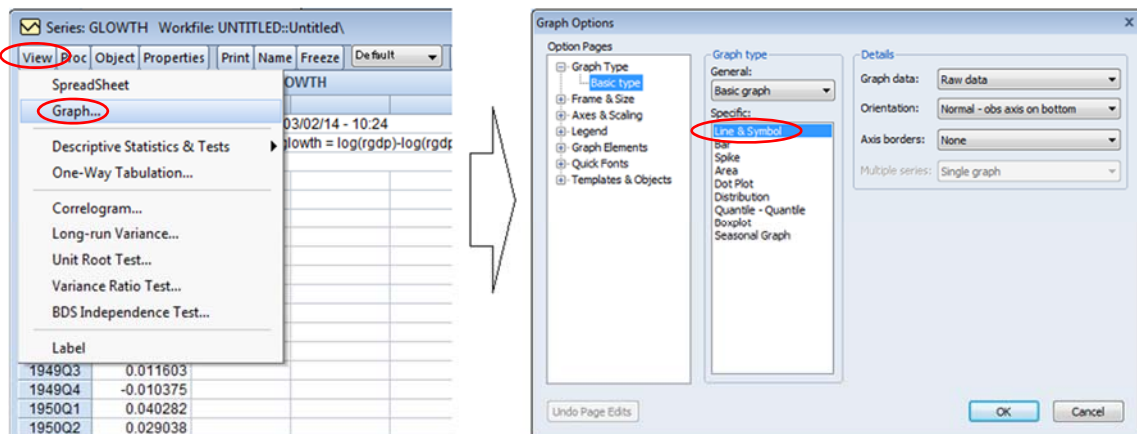
これまで見てきたように、分析の際に、あらかじめ変数を加工することで適切な分析ができる。ただし、推定時点で関数を用いて推定式を表現することも可能であり、全ての変数をあらかじめ定義する必要はない。

2.2 グラフ

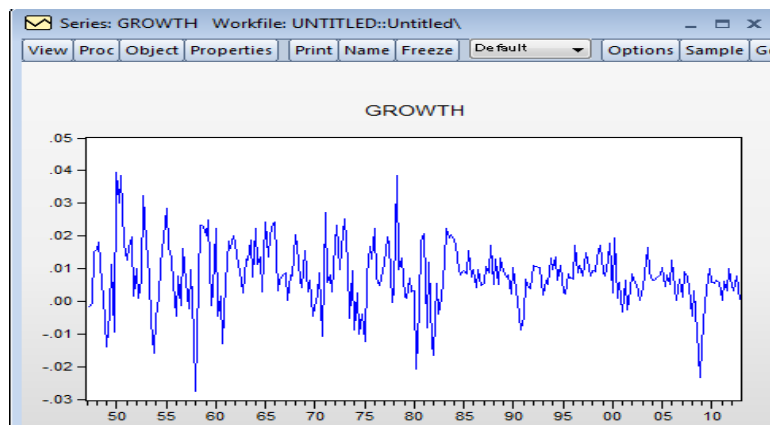
変数のグラフを図示してみよう。Workfile ウィンドウから、系列 growth を選び、ダブルクリックすると、系列の詳細を表示する Series ウィンドウが出てくる。1947Q1 については NA とある。これは 1947Q1 の成長率を計算するには、1946Q4 の GDP が必要であるためである。

Year	Growth Rate
1947Q1	NA
1947Q2	-0.001526
1947Q3	-0.000849
1947Q4	0.015057
1948Q1	0.015768
1948Q2	0.019221
1948Q3	0.005375
1948Q4	0.001554
1949Q1	-0.014015
1949Q2	-0.003644
1949Q3	0.011161
1949Q4	-0.009419
1950Q1	0.039614

次に、Series ウィンドウのメニューバーから、「View」→「Graph」を選択すると（左下図）、Graph Options のウィンドウが表示される（右下図）。ここではグラフの種類を指定する。graph type では折れ線グラフ (Line & Symbol) や棒グラフ (Bar) など、グラフの種類を選択する。ここでは「Line & Symbol」を選択しよう。



OK を押すと、グラフが Series ウィンドウに表示される（下図参照）。こうした図をみることによって、構造変化や外れ値の存在を確認できる。分析の前には、各系列を図示して、どのようなデータかを考える習慣を身に付けてもらいたい。



3. 回帰分析

データが準備できたら、実際に回帰分析を行ってみよう。線形モデルの最小2乗(Ordinary Least Squares、略してOLS)推定では、コマンドウィンドウに

ls 被説明変数 説明変数1 説明変数2 ...

と入力すればよい。たとえば、Yを被説明変数、X1とX2を説明変数とし、定数項cを含めたモデル ($Y = c + \beta_1 X_1 + \beta_2 X_2$) を推定したいなら、コマンドウィンドウに

ls Y c X1 X2

と書けばよい。ここで、Enterを押すと、推定結果が表示される。もし定数項をいれたくないなら、cを除いて入力すればよい。

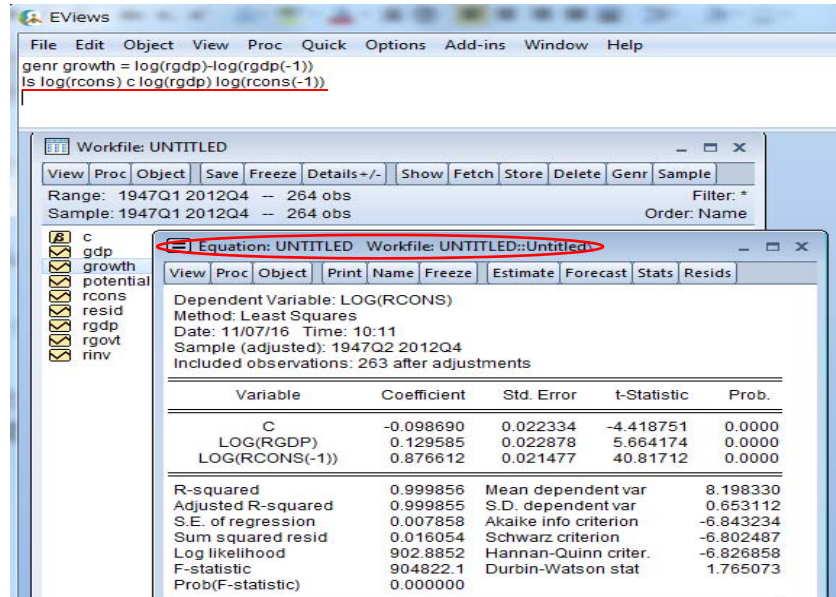
ls Y X1 X2

3.1 消費関数の推定

実際の分析を通じて理解を深めよう。被説明変数を消費額(RCONS)、説明変数を所得(RGDP)、前期の消費額(RCONS(-1))とした消費関数を推定する。これは

ls log(rcons) c log(rgdp) log(rcons(-1))

と入力し、Enterをおすと、下図のようなEquationウィンドウで推定結果が表示される。なお、全ての変数は対数表示とした。ここでは、事前に変数を対数変換していなかったの、推定時点で対数であることを明示している。



下表では、それぞれの情報が何を意味しているかをまとめている。

項目	説明
Coefficient	係数の推定値
Std. Error	係数の標準誤差
t-Statistic	t値(帰無仮説: 係数=0)

F-Statistic	定数項以外のすべての係数が 0 という仮説を検定した F 値
Prob.	p 値
R-squared	決定係数 (R^2)
Adjusted R-squared	自由度調整済み決定係数 (\bar{R}^2)

係数の推定値を標準誤差で割ったものが、t 値となっている。また、対応する p 値をみることで、何%の有意水準で有意な結果なのかを判断できる。F 値は、定数項以外のすべての係数が 0 であるかを検定しており、この仮説が棄却できれば、どれかの変数は Y の動きを説明するうえで意味があるといえる。

この推定結果を簡潔に表すと、

$$\log(rcons_t) = -0.099 + 0.130 \log(rgdp_t) + 0.877 \log(rcons_{t-1})$$

(-4.42) (5.66) (40.82)

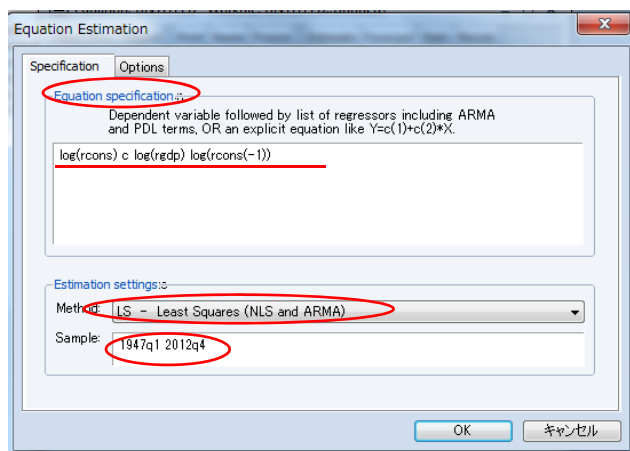
$$\bar{R}^2 = 0.99$$

となる（カッコ内は t 値を表す）。各係数の t 値は大きな値をとっていることから、係数はゼロと有意に異なる。また、自由度調整済み決定係数 \bar{R}^2 が 1 に近いことから、当てはまりのよいモデルといえる。時系列データを扱うと、決定係数は高くなる傾向があり、クロスセクションデータを扱うと、決定係数が低くなる傾向がある。

こうした推定は、メニューバーの「Quick」→「Estimate Equation」を選択して Equation Estimation ウィンドウを開き、Equation specification に

$$\log(rcons) \text{ c } \log(rgdp) \log(rcons(-1))$$

と入力してもよい（下図参照）。デフォルト設定として、Method（推定方法）は最小 2 乗法 LS、Sample（分析期間）は全期間「1947Q1 2012Q4」と入力されている（つまり、1947 年第一四半期から 2012 年第四四半期まで）。



この方法では、期間を明示的に選択できるため、Sample を変更することで部分期間だけの推定も可能である。さまざまな部分期間を推定することで、推定値が時間を通じて安定

しているかを調べることができる。たとえば、1990年Q1から2010Q4まで分析したいなら、Sampleを「1990Q1 2010Q4」とすればよい。

ちなみに、Equation specificationには数式を直接入力してもよい。たとえば、

$$\log(\text{rcons}) = c(1) + c(2) \log(\text{rgdp}) + c(3) \log(\text{rcons}(-1))$$

の代わりに、

$$\log(\text{rcons}) = c(1) + c(2) * \log(\text{rgdp}) + c(3) * \log(\text{rcons}(-1))$$

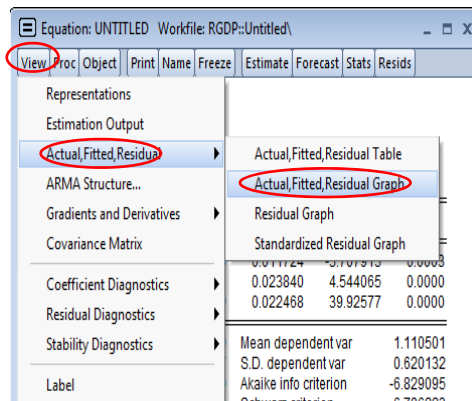
と入力する。ここでc(1)、c(2)、c(3)は推定するパラメータを意味する。そして、OKを入力すると同じ結果が得られる。線形モデルを推定する限り、数式を直接入力することのメリットはない。しかし、非線形モデルの推定だと数式を直接入力することが必要となってくる。たとえば、GDPと消費との関係は非線形と考えて、

$$\log(\text{rcons}) = c(1) + c(2) * \log(\text{rgdp})^c(3) + c(4) * \log(\text{rcons}(-1))$$

というモデルを考えることもできる。つまり、この式では、 $\log(\text{rgdp})$ のc(3)乗が説明変数になっており、c(3)も未知のパラメータになっている。

3.2 モデルの理論値と残差

モデルの理論値や残差が知りたい場合、Seriesウィンドウのメニューバーから、「View」→「Actual, Fitted, Residual」→「Actual, Fitted, Residual Graph」を選択しよう。



すると、Seriesウィンドウで左下図のようなグラフが表示される。青線が残差、赤線が実際の値、緑線が理論値を表している（この場合、モデルの当てはまりが良いため、赤線と緑線はほぼ一致）。なお、各系列の数値を見たい場合は、「View」→「Actual, Fitted, Residual」→「Actual, Fitted, Residual Table」を選択して系列を表示しよう（右下図）。右下図の左側は、各系列の数値（ActualはYの値、FittedはYの理論値、Residualは残差）を、右側は残差の動きを図示したものである。この図を詳しくみることで、残差に系列相関があるか、当てはまりの悪い時期（外れ値？）があるかを視覚的に調べることができる。もし系列相関があるなら、説明変数を追加することで、こうした系統的な変動をとらえることができる。また、当てはまりの悪い時期があれば、モデルではうまく説明できていない部分であり、何らかの変数（ダミー変数など）を追加するかを検討してもよいだろう。

この場合、残差に正の系列相関（残差がプラスなら次期にもプラスになりやすい）がありそうである。系列相関をとるために、1期前の消費だけでなく、2期前、3期前の消費も説明変数として加えることを検討してもよいかもしれない。

