2023/05/24

藪友良

第 7 章の答え

練習問題 1

7.1.2 節で述べた通り、個別の t 検定で結合仮説を検定すると有意水準を適切に設定することが困難となる。このため、F 検定を用いて同時検定をすることが望ましい。

練習問題 2

(a) 帰無仮説H_0と対立仮説H_1は、それぞれ次のように設定すればよい。

$$H_0\colon\ \beta_2 = 0、\beta_3 = 0$$

$$H_1\colon\ 帰無仮説H_0は誤り$$

ここで、除外制約の数は計$q = 2$となる。

(b) 帰無仮説H_0が正しいもとで、モデルは次のようになる。

$$Y = \alpha + \beta_1 X_1 + u$$

(c) $n = 50$、$K = 3$、$q = 2$、$SSR_0 = 150$、$SSR_1 = 100$を F 統計量の式に代入すると、F 値は次のようになる。

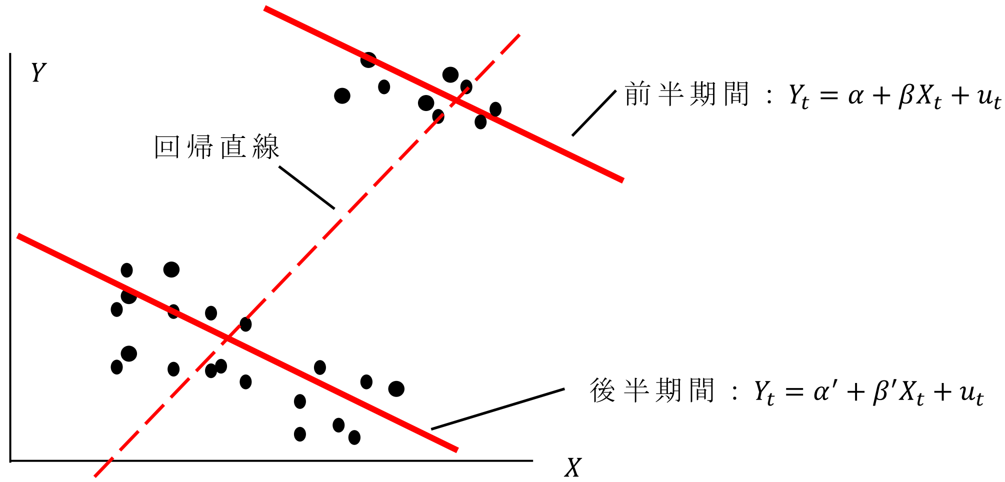
$$F = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(n - K - 1)} = \frac{(150 - 100)/2}{100/(50 - 3 - 1)} = 23 \times 0.5 = 11.5$$

帰無仮説H_0が正しいもとで、F 統計量は F 分布(自由度 2、46)に従う。このとき、有意水準 5%の臨界値は、Excel で「=FINV(0.05,2,46)」と入力すれば 3.199582 と分かる。F 値は 11.5 であり、これは臨界値 3.199582 を上回るため、帰無仮説H_0は棄却される。つまり、「説明変数X_2とX_3に説明力がない」とはいえない。

練習問題 3

様々な原因が考えられるが、ここでは 1 つの可能性を考える。下図の実線は、XとYの真の関係を表している。前半期間と後半期間とも、Xの係数は負であるが($\beta < 0$、$\beta' < 0$)、後半期間において定数項は小さくなっている($\alpha' < \alpha$)。このと

き、パラメータに生じた構造変化を考慮しないで、すべてのデータをまとめて推定すると、回帰直線は点線のようになり、回帰直線の傾きは正となる。



この例から、構造変化を考慮しないでモデルを推定してしまうと、推定結果にバイアスを生じさせることが理解できる。

練習問題 4

男女のモデルを統合した次式を考える。

$$Y_i = \alpha + \beta X_i + \theta_0 F_i + \theta_1 F_i X_i + u_i$$

ここで、F_iは女性ダミーであり、パラメータθ_0とθ_1は次のように定義される。

$$\theta_0 = \alpha' - \alpha、\theta_1 = \beta' - \beta$$

たとえば、iが男性なら$F_i = F_i X_i = 0$となるので、男性のモデルは、

$$Y_i = \alpha + \beta X_i + u_i$$

となる。iが女性なら$F_i = 1$、$F_i X_i = X_i$となるので、女性のモデルは、

$$Y_i = \alpha' + \beta' X_i + u_i$$

となる。ここで、帰無仮説と対立仮説を次のように設定し F 検定をすれば、男女でパラメータが同じであるかを検証できる。

$$H_0: \theta_0 = 0, \theta_1 = 0$$

H_1: H_0は誤りである

練習問題 5

対立仮説H_1が正しいとしたモデルでは、残差 2 乗和は$SSR_1 = \sum_{i=1}^{n} \hat{u}_{1i}^2$となる。

また、決定係数R_1^2は、次のようになる。

$$R_1^2 = 1 - \frac{SSR_1}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

つまり、残差 2 乗和SSR_1は、決定係数R_1^2を用いて、次のように表現できる。

$$SSR_1 = (1 - R_1^2)\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

同様に、帰無仮説H_0が正しいとしたモデルからの残差 2 乗和は次のようになる。

$$SSR_0 = (1 - R_0^2)\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

これらを F 統計量の式に代入すると、次のようになる。

$$F = \frac{(SSR_0 - SSR_1)/q}{SSR_1/(n-K-1)} = \frac{((1-R_0^2)-(1-R_1^2))\sum_{i=1}^{n}(Y_i - \bar{Y})^2/q}{(1-R_1^2)\sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-K-1)} = \frac{(R_1^2 - R_0^2)/q}{(1-R_1^2)/(n-K-1)}$$

つまり、対立仮説H_1が正しいとしたモデルからの決定係数R_1^2が、帰無仮説H_0が正しいとしたモデルからの決定係数R_0^2より高くなると、F 統計量の値は大きくなる。

練習問題 6

帰無仮説H_0が正しいなら、モデルは、次のようになる。

$$Y_i = \alpha + u_i$$

このとき、OLS 推定量は$\hat{\alpha} = \bar{Y}$となるため、残差は$\hat{u}_{0i} = Y_i - \hat{\alpha} = Y_i - \bar{Y}$となる(2 章の練習問題 7 参照)。つまり、残差 2 乗和SSR_0は、Y_iの偏差 2 乗和となる。

$$SSR_0 = \sum_{i=1}^{n}\hat{u}_{0i}^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

このとき、決定係数R_0^2は、次のようになる。

$$R_0^2 = 1 - \frac{SSR_0}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = 0$$

練習問題 6 の結果に、$R_0^2 = 0$を代入すると、次のようになる。

$$F = \frac{R_1^2/K}{(1-R_1^2)/(n-K-1)}$$

この式から、R_1^2が 1 に近づくと、F値が∞に発散することがわかる。つまり、対

立仮説H_1におけるモデルを推定し、説明変数の当てはまりが良ければ、帰無仮説H_0を棄却できる。

練習問題 7

$T=200$である場合、構造変化点の候補の始期(T_{min})と終期(T_{max})は、

$$T_{min}=0.15\times 200=30$$

$$T_{max}=(1-0.15)\times 200=170$$

となる。したがって、構造変化点の候補T_Bは次のとおりである。

30、31、32、…、168、169、170

$T=1000$である場合、構造変化点の候補の始期(T_{min})と終期(T_{max})は、

$$T_{min}=0.15\times 1000=150$$

$$T_{max}=(1-0.15)\times 1000=850$$

となる。したがって、構造変化点の候補T_Bは次のとおりである。

150、151、152、…、848、849、850

練習問題 8

説明変数が 2 個の場合、定数項を含めると排除制約の数は$q=3$となる。よって、臨界値は有意水準 10%なら 4.09、5%なら 4.71、1%なら 6.02 である。sup F = 4.50 は、有意水準 10%の臨界値 4.09 を上回るため、有意水準 10%で帰無仮説は棄却される。つまり、パラメータに構造変化がない、とはいえない。

練習問題 9

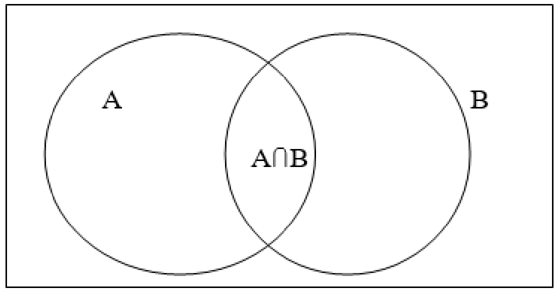
統計学で学習する加法定理では、事象AまたはBが生じる確率$P\{A\cup B\}$は、

$$P\{A\cup B\}=P\{A\}+P\{B\}-P\{A\cap B\}$$

となる。つまり、事象 A または B が生じる確率は、A の確率と B の確率を足したのち、A と B が同時に生じる確率($P\{A\cap B\}$)を引いたものとなる。

下図は、ベン図を用いてこれらの事象を図式化している。長方形 □ で囲まれた領域内は標本空間であり、それぞれ〇で囲まれた領域は事象 A と B となる。事象 A と B が重なる部分は $A\cap B$ である。図をみると、$P\{A\cup B\}$を求めるために、$P\{A\}$と$P\{B\}$の和を求めると、A と B の共通部分($P\{A\cap B\}$)が 2 回分も含ま

れてしまうので、その和から余分な 1 回分($P\{A \cap B\}$)を引く必要があると理解できる。これが加法定理である。



確率の公理を用いた加法定理の証明は、藪友良「入門　実践する統計学」(東洋経済新報社、2012 年)の 4 章補足を参照されたい。

練習問題 10

構造変化点をT_Bとした F 統計量$F(T_B)$は、次のように表現できる。

$$F(T_B) = \frac{(SSR_0 - SSR_1)/(K+1)}{SSR_1/(T-2(K+1))} = \frac{T-2(K+1)}{K+1}\left(\frac{SSR_0}{SSR_1} - 1\right)$$

ここでは、$F(T_B)$の構成要素において、構造変化点T_Bに依存しているのは残差 2 乗和SSR_1のみであることを示す。

まず、残差 2 乗和SSR_0は、帰無仮説H_0(構造変化なし)が正しい前提で、(7)式、つまり、下式を推定することで得られる。

$$Y_t = \alpha + \beta_1 X_{1t} + \ldots + \beta_K X_{Kt} + u_t$$

このため、残差 2 乗和SSR_0は、どのT_Bを用いても同じ値となる。

次に、残差 2 乗和SSR_1は、対立仮説H_1(構造変化あり)が正しい前提で、(6)式、つまり、下式を推定することで得られる。

$$Y_t = \alpha + \beta_1 X_{1,t} + \cdots + \beta_K X_{K,t} + \theta_0 D_t + \theta_1 D_t X_{1,t} + \ldots + \theta_K D_t X_{K,t} + u_t$$

D_tは時点tが前半期間(1、2、⋯、T_B)ならば 0 をとり、後半期間(T_B+1、T_B+2、⋯、T)ならば 1 をとる。D_tはT_Bの選択によって値が変わるため、残差 2 乗和SSR_1もT_Bの選択によって値が変わる。

以上から、$F(T_B)$の構成要素のうち、T_Bに依存しているのは残差 2 乗和SSR_1のみである。残差 2 乗和SSR_1が小さいほど$F(T_B)$が大きくなることに注意すると、$F(T_B)$を最大にするT_Bとは、(6)式の残差 2 乗和SSR_1を最小にするT_Bに他ならない。

練習問題 11

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。

STATA の再現コード

```
**　データの読み込み
use stockprice_data.dta
tsset time


**　7.3.3 節の推定
reg ds ds_1 ds_2 ds_3 s_ma
```

データを読み込んだ後、データが時系列データであることを認識させるため、tsset time としよう。そして、通常の OLS 推定をすれば次の推定結果が得られる。

```
      Source |       SS           df       MS      Number of obs   =     6,677
-------------+----------------------------------   F(4, 6672)      =      3.33
       Model |  29.8494008         4  7.46235019   Prob > F        =    0.0098
    Residual |  14940.9234     6,672  2.23934704   R-squared       =    0.0020
-------------+----------------------------------   Adj R-squared   =    0.0014
       Total |  14970.7728     6,676  2.24247646   Root MSE        =    1.4964

------------------------------------------------------------------------------
          ds |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        ds_1 |  -.0346764   .0130919    -2.65   0.008    -.0603406   -.0090122
        ds_2 |  -.0234359   .0130094    -1.80   0.072    -.0489385    .0020666
        ds_3 |  -.0035702   .0129095    -0.28   0.782    -.0288771    .0217366
        s_ma |  -.0023591    .004883    -0.48   0.629    -.0119314    .0072131
       _cons |  -.0013814    .018316    -0.08   0.940    -.0372865    .0345238
------------------------------------------------------------------------------
```

赤枠内をみると、F 値は 3.33 であるとわかる。これはすべての係数が 0 とした帰無仮説に対応した F 値である。また、Prob>F は 0.0098 とあるが、これは F 検定の p 値に対応している。p 値が 0.01 を下回るため、すべての係数が 0 とした帰無仮説は有意水準 1%で棄却される。

練習問題 12

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
**　データの読み込み
```

```
use growth_data.dta
gen time = q(1981q1)+_n-1
tsset time, quarterly

**　変数の定義
gen d1991 = (time>q(1991Q1))
gen d1991growth = d1991*l.growth

**7.4.1 節の推定結果
reg growth l.growth
reg growth l.growth d1991 d1991growth
test d1991 d1991growth

** 同じ検定は estat を用いてもできる
reg growth l.growth
estat sbknown, break(tq(1991Q2))

** 7.4.2 節の推定結果
reg growth l.growth
estat sbsingle, trim(15)
```

① データ growth_data.dta を読み込もう。

```
use growth_data.dta
```

② 時系列データであることを認識させるため、変数 time を定義する。

```
gen time = q(1981q1)+_n-1
tsset time, quarterly
```

データは四半期データであり、1981 年第 1 四半期から始まる。

③ ダミー変数と交差項を次のように定義する。

```
gen d1991 = (time>q(1991Q1))
gen d1991growth = d1991*l.growth
```

d1991 は、1991 年第 1 四半期までは 0、それ以降は 1 となるダミー変数である。

d1991growth は、1 期前の成長率 l.growth とダミー変数との交差項となる。

④ 帰無仮説(構造変化なし)が正しいという前提で推定すると、次のようになる。

reg growth l.growth

```
. reg growth l.growth

      Source |       SS           df       MS      Number of obs   =       107
-------------+----------------------------------   F(1, 105)       =    234.52
       Model |  354.515149         1  354.515149   Prob > F        =    0.0000
    Residual |  158.721217       105  1.51163064   R-squared       =    0.6907
-------------+----------------------------------   Adj R-squared   =    0.6878
       Total |  513.236365       106   4.8418525   Root MSE        =    1.2295

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      growth |
         L1. |   .8304376   .0542266    15.31   0.000     .7229163     .937959
             |
       _cons |   .3965092   .1812967     2.19   0.031     .0370313    .7559871
------------------------------------------------------------------------------
```

左上の赤枠が残差 2 乗和であり、これは 158.72 とわかる。次に、対立仮説(構造変化あり)が正しい前提で推定すると、次のようになる。

reg growth l.growth d1991 d1991growth

```
. reg growth l.growth d1991 d1991growth

      Source |       SS           df       MS      Number of obs   =       107
-------------+----------------------------------   F(3, 103)       =     96.64
       Model |  378.692668         3  126.230889   Prob > F        =    0.0000
    Residual |  134.543698       103  1.30624949   R-squared       =    0.7379
-------------+----------------------------------   Adj R-squared   =    0.7302
       Total |  513.236365       106   4.8418525   Root MSE        =    1.1429

------------------------------------------------------------------------------
      growth |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      growth |
         L1. |   .5040964   .1216393     4.14   0.000     .2628536    .7453392
             |
       d1991 |   -1.94702   .6105075    -3.19   0.002    -3.157818   -.7362225
 d1991growth |   .1735425    .151043     1.15   0.253    -.1260156    .4731006
       _cons |   2.296806   .5825721     3.94   0.000     1.141412      3.4522
------------------------------------------------------------------------------
```

左上の赤枠が残差 2 乗和であり、これは 134.54 とわかる。残差 2 乗和を用いて、F 検定ができるが、面倒なので、コマンドを使って行う。

ここで、d1991 と d1991growth の係数が 0 とした F 検定は、

test d1991 d1991growth

とすればよい。下画面をみると、F 値は 9.25 である。Prob>F は p 値であり、これは 0.0002 と 1%を下回り、有意水準 1%で帰無仮説(構造変化なし)を棄却できる。

```
. test d1991 d1991growth

 ( 1)  d1991 = 0
 ( 2)  d1991growth = 0

       F(  2,   103) =    9.25
            Prob > F =    0.0002
```

同じ推定は、次のコマンドでも実行できる。

reg growth l.growth

estat sbknown, break(tq(1991Q2))

sbknown は構造変化日が既知であることを示す。また、break(tq(1991Q2))は構造変化日が 1991q2 であることを示す。STATA では、構造変化日は T_B+1 として定義されており、構造変化日は 1991Q1 ではなく、1991Q2 とした。推定結果は、下画面のようになる。

```
Wald test for a structural break: Known break date

                                  Number of obs   =        107
Sample:       1981q2 - 2007q4
Break date:   1991q2
Ho: No structural break

              chi2(2)       =    18.5091
              Prob > chi2   =     0.0001

Exogenous variables:          L.growth
Coefficients included in test: L.growth _cons
```

ワルド統計量(Wald statistic)が 18.5091 と計算されており、教科書では紹介していないが、

ワルド統計量 ＝F 統計量×制約数 q

という関係がある。つまり、

F 統計量 ＝ ワルド統計量÷制約数 q

という関係になる。この場合、制約数 q は 2 であるため、F 統計量は 18.5091÷2=9.25 である。

⑤ 構造変化日が未知であるとき、次のようにすればよい。

reg growth l.growth

estat sbsingle, trim(15)

sbsingle は、構造変化が 1 度だけであるが、未知であることを示す。trim(15)は、最初と最後の 15%のデータを、構造変化の候補から除くことを示す。推定結果は以下

のとおり。構造変化日の推定値 Estimated break date は 1991q2 となる。STATA では、構造変化日は T_B+1 として定義されており、教科書の定義に合わせると、構造変化日は 1991 年第 1 四半期であることを意味している。また、ワルド統計量は 18.5091 であり、p 値は 0.0023 と 1%を下回っている。

```
Test for a structural break: Unknown break date

                              Number of obs =        107

Full sample:                  1981q2 - 2007q4
Trimmed sample:               1985q3 - 2003q4
Estimated break date:         1991q2
Ho: No structural break

     Test           Statistic         p-value
-----------------------------------------------
     swald            18.5091          0.0023
-----------------------------------------------
Exogenous variables:            L.growth
Coefficients included in test: L.growth _cons
```

練習問題 13

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
use dailyint_data.dta
gen time = _n
tsset time
gen ds = 100*ln(spot/l.spot)
gen intj = intervention/10000


reg ds intj
estat sbsingle, trim(15)
reg ds intj if(time<=1056)
reg ds intj if(time>1056)


**推定結果の頑健性を調べる
```

```
reg ds intj l.ds

estat sbsingle, trim(15)

reg ds intj l.ds if(time<=1071)

reg ds intj l.ds if(time>1071)
```

① データ dailyint_data.dta を読み込もう。

```
use dailyint_data.dta
```

そして、時系列データであることを認識させる。

```
gen time = _n

tsset time
```

② データを定義する。ドル円レート spot の対数差として、為替変化率 ds を定義する。また、介入額は億円単位なので、10000 で割ることで兆円単位にする。

```
gen ds = 100*(ln(spot)-ln(l.spot))

gen intj = intervention/10000
```

③ 構造変化日が未知なので、次のようにして sup F 検定をしよう。

```
reg ds intj

estat sbsingle, trim(15)
```

そうすると、以下の画面が出力される。

```
Test for a structural break: Unknown break date

                              Number of obs =      3,054

Full sample:                  2    - 3055
Trimmed sample:               461 - 2597
Estimated break date:         1057
Ho: No structural break

     Test                Statistic            p-value

     swald                33.8990              0.0000

Exogenous variables:            intj
Coefficients included in test: intj _cons
```

ワルド統計量は 33.8990 であり、これを 2 で割ると、sup F は 16.945 となる。構造変化日は 1057(1995/4/19 に当たる)となっているので、教科書の構造変化日 T_B は 1056(1995/4/18)となる。

④ 前半と後半に分けて別々に推定するには、次のようにすればよい。

```
reg ds intj if(time<=1056)
```

reg ds intj if(time>1056)

if(time<=1056)は、1995/4/18 までのデータを用いた推定であり、if(time>1056)は、1995/4/19 からのデータを用いた推定である。

⑤ 説明変数に、1 期前の為替変化率をいれて構造変化の検定をするには、次のように入力すればよい。

```
reg ds intj l.ds
estat sbsingle, trim(15)
```

そうすると、以下の画面が出力される。

```
Test for a structural break: Unknown break date

                              Number of obs =      3,053

Full sample:                  3   - 3055
Trimmed sample:               461 - 2598
Estimated break date:         1072
Ho: No structural break

     Test             Statistic          p-value

     swald             39.4350           0.0000

Exogenous variables:            intj L.ds
Coefficients included in test: intj L.ds _cons
```

構造変化日は 1072(1995/5/10)となっているので、教科書の構造変化日の定義である T_B は 1071(1995/5/9)となる。ワルド統計量は 39.435 であり、これを 2 で割ることで、sup F は 19.7175 とわかる。この結果から、1 期前の為替変化率を入れても、構造変化があること、その時点は、ほぼ同じであることがわかる。

練習問題 14

ウェブサイトから、再現に必要なデータと STATA の do file をダウンロードできる。STATA 再現コードは以下のとおり。

STATA の再現コード

```
use monthlyint_data.dta
gen time = m(1991m4)+_n-1
tsset time, monthly
replace intj = intj/10000
```

```
replace intj_gaika = intj_gaika/10000
replace intj_taimin = intj_taimin/10000


reg intj intj_gaika   /*推定結果(外貨準備)*/
test _b[intj_gaika]=1
test _cons=0, accumulate
reg intj intj_taimin /*推定結果(対民収支)*/
test _b[intj_taimin]=1
test _cons=0, accumulate
```

① データ monthlyint_data.dta を読み込もう。

```
use monthlyint_data.dta
```

そして、1991 年 4 月から始まる時系列データであることを認識させる。

```
gen time = m(1991m4)+_n-1
tsset time, monthly
```

② 変数が億円単位なので、すべて兆円単位に変換する。

```
replace intj = intj/10000
replace intj_gaika = intj_gaika/10000
replace intj_taimin = intj_taimin/10000
```

ここで、replace は変数を置き換えるためのコマンド。こうすれば、同じ変数名が使える。

③ 外貨準備について、H_0: $\alpha = 0$, $\beta = 1$とした F 検定をしよう。これは、次のようにすればよい。

```
reg intj intj_gaika
test _b[intj_gaika]=1
test _cons=0, accumulate
```

つまり、_b[intj_gaika]=1 が$\beta = 1$、_cons=0 が$\alpha = 0$に対応している。F 値は 1.11 であり、p 値は 0.3332 と 10%を上回るので、有意な結果ではない。

```
 ( 1)  intj_gaika = 1
 ( 2)  _cons = 0

       F(  2,    94) =    1.11
            Prob > F =    0.3332
```

同様にして、対民収支についても、H_0: $\alpha = 0,\ \beta = 1$とした F 検定をしよう。これは次のようにすればよい。

```
reg intj intj_taimin /*推定結果(対民収支)*/
test _b[intj_taimin]=1
test _cons=0, accumulate
```

推定結果については自分で確認してみてほしい。