宏观与微观数据的混合回归及 Stata 应用：mixregress

王群勇（南开大学数量经济研究所, QunyongWang@outlook.com)

内容

混合回归

Stata 混合回归: mixregress

混频回归

低频时间序列 vs 高频时间序列（midasreg）

宏观时间序列 vs 微观面板（mixregress）

混合回归

混合回归模型：

$$y_t = z_t\beta + \alpha \sum_{i=1}^{n_t} w_{it}\, x_{it} + \epsilon_t.$$

其中，(y_t, z_t) $(t = 1, \dots, T)$ 为宏观变量，x_{it}为微观变量，i表示第i个体（比如个人、企业等）。w_{it}表示个体i的权重。即微观x_{it}的加权和来解释和预测y_t。

n_t: t期有n_t个微观个体，微观面板数据可以是平衡或非平衡的。

设有m个微观指标，权重函数设为

$$w_{it} = \theta_1 w_{1,it} + \cdots + \theta_m w_{m,it}.$$

混合回归

要识别θ、α，必须施加必要的约束。Gysel(2019)用企业收益预测季度 GDP 的模型

$$GDP_{t+1} = \alpha + \beta_{earn} \sum_{i=1}^{n_t} w_{it}\, earn_{it} + \sum \beta_k z_{k,t} + \epsilon_{t+1}.$$

$$w_{it} = (\theta_1 ew_{it} + \theta_2 vw_{it} + \tau \times smth_{it}.$$

Gysel (2019)约束$\theta_1 + \theta_2 = 1$, 即

$$w_{it} = (1-\theta) ew_{it} + \theta vw_{it} + \tau \times smth_{it}.$$

其中，$(1-\theta) ew_{it} + \theta vw_{it}$叫做基准成分(benchmark component), $\tau \times smth_{it}$叫做调节成分（tilt component）.

其它约束：$\beta_{earn} = 1$，或者$\tau = 1$.

估计方法

非线性最小二乘法：$y_t = x_t(\beta) + u_t$，最小化残差平方和：

$$SSR(\beta) = \sum_{t=1}^{T} \left(y_t - x_t(\beta)\right)^2.$$

数值优化算法：Newton-Raphson, BHHH 等。

同方差条件下，NLS 估计量$\hat{\beta}$的方差为

$$\widehat{Var}(\hat{\beta}) = s^2 \left(\hat{X}'\hat{X}\right)^{-1}.$$

其中，$X(\beta) = \partial x(\beta)/\partial\beta$, $\hat{X} = \partial x(\hat{\beta})/\partial\hat{\beta}$.

$$s^2 = \frac{1}{T-k} \sum_{t=1}^{T} \hat{u}_t^2.$$

估计方法

异方差稳健方差：

$$\widehat{Var}(\hat{\beta}) = \left(\hat{X}'\hat{X}\right)^{-1} \left(\hat{X}'\hat{\Omega}\hat{X}\right) \left(\hat{X}'\hat{X}\right)^{-1},$$

其中，$\hat{\Omega}$为对角矩阵，元素为$\hat{u}_t^2$，

$$\hat{X}'\hat{\Omega}\hat{X} = \hat{X}'\hat{\Omega}_0\hat{X} = \frac{T}{T-k}\sum_{t=1}^{T}\hat{u}_t^2\, x_t{}'x_t$$

Newey-West (1987)异方差自相关稳健方差

$$\hat{X}'\hat{\Omega}\hat{X} = \hat{X}'\hat{\Omega}_0\hat{X} + \frac{n}{n-k}\sum_{l=1}^{m}\left(1-\frac{l}{m+1}\right)\sum_{t=l+1}^{T}\hat{u}_t\,\hat{u}_{t-l}(x_t{}'x_{t-l} + x_{t-l}{}'x_t).$$

Stata

model: $y_t = z_t\beta + \alpha\sum_{i=1}^{n_t} w_{it}\, x_{it} + \epsilon_t$.

`mixregress` *depvar indeps* [*if*] [*in*], `xmicro(` *varname* `)` `wmicro(` *varlist* `)` `idtype(` *integer* `)` `vce(` *string* `)` `maxlag(` *integer* `)` [*options*]

`xmicro(` *varname* `)`： 微观变量z_{it}

`wmicro(` *varlist* `)`： 微观变量w_{it}

`idtype(` *integer* `)`： 识别策略。`idtype(0)`: 约束$\theta_1 + \theta_2 = 1$, `idtype(1)`约束$\theta_2 = 1$.

`vce(` *string* `)`： `vce(0)`（同方差）; `vce(1)`： 异方差稳健； `vce(2)`： 异方差自相关稳健

`maxlag(` *integer* `)`： Newey-West (1987)标准差的滞后阶数。默认值为ceil$\left(0.43T^{1/3}\right)$ (Stock and Watson , 2011).

例

```
. use "macro2.dta", clear

. tsset

Time variable: yearquarter, 2003q1 to 2019q4
        Delta: 1 quarter

. des ppi m0

Variable      Storage   Display    Value
    name         type    format    label      Variable label
--------------------------------------------------------------------------------

----------------------------------
ppi             float   %9.0g
```

```
m0              float   %9.0g

.
. use "micro2.dta", clear

. xtset

Panel variable: stock (unbalanced)
 Time variable: yearquarter, 2003q1 to 2019q4, but with gaps
         Delta: 1 quarter

. des earn ew vw SA_w

Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------
----------------------------------

earn            float   %9.0g                 季度盈余同比增长率
ew              float   %9.0g
vw              float   %9.0g
SA_w            float   %9.0g
```

例

```
. use "micro2.dta", clear

. qui merge m:1 yearquarter using macro2

. mixregress ppi Lppi m0, xmicro(earn) wmicro(ew vw) vce(2) nolog

Log-likelihood =       182.4143              Number of obs    =       107425
AIC            =      -354.8285              Number of group  =           67
BIC            =      -343.8051              R-squared        =       0.8591
HQIC           =      -350.4665              Adj R-squared    =       0.8475
DW             =         1.5174              Root MSE         =       0.0167
------------------------------------------------------------------------------
             |                HAC
         ppi | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
macro        |
        Lppi |   .9004863   .0386545    23.30   0.000      .824725    .9762477
          m0 |  -.0003632   .0001122    -3.24   0.001     -.000583   -.0001433
       Wearn |   .0987397   .0216344     4.56   0.000      .056337    .1411424
       _cons |  -.0209212   .0050621    -4.13   0.000    -.0308428   -.0109997
-------------+----------------------------------------------------------------
micro        |
```

```
          ew |   .8944858   .5411519     1.65   0.098    -.1661525    1.955124
          vw |   .1055142   .5411519     0.19   0.845    -.9551241    1.166152
-------------+----------------------------------------------------------------
variance     |
     lnsigma |  -4.141539   .0007202 -5750.81   0.000    -4.142951   -4.140128
------------------------------------------------------------------------------

. est store mod
```

例

```
. egen earnsum = total(earn), by(yearquarter)

. reg ppi Lppi m0 earnsum if vw<=0.006

      Source |       SS           df       MS      Number of obs   =   106,454
-------------+----------------------------------   F(3, 106450)    >  99999.00
       Model |  157.013825         3  52.3379417   Prob > F        =    0.0000
    Residual |  29.3202764   106,450  .000275437   R-squared       =    0.8426
-------------+----------------------------------   Adj R-squared   =    0.8426
       Total |  186.334102   106,453  .001750388   Root MSE        =     .0166

------------------------------------------------------------------------------
         ppi | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        Lppi |   .8885288   .0012488   711.50   0.000     .8860812    .8909765
          m0 |  -.0005662   .0000112   -50.38   0.000    -.0005882   -.0005441
     earnsum |   .0000459   2.06e-07   222.85   0.000     .0000455    .0000463
       _cons |  -.0172298   .0000954  -180.59   0.000    -.0174168   -.0170428
------------------------------------------------------------------------------
```

例

```
. mixregress ppi Lppi m0 if vw<=0.006, xmicro(earn) wmicro(ew vw) vce(2) nolog

Log-likelihood =        182.9834                Number of obs   =       106454
AIC            =       -355.9667                Number of group =           67
BIC            =       -344.9433                R-squared       =       0.8614
HQIC           =       -351.6047                Adj R-squared   =       0.8501
DW             =          1.5601                Root MSE        =       0.0165
------------------------------------------------------------------------------
```

ppi	Coefficient	HAC std. err.	z	P>|z|	[95% conf.	interval]
macro						
Lppi	.8923568	.0390872	22.83	0.000	.8157474	.9689663
m0	-.0003377	.000119	-2.84	0.005	-.0005709	-.0001044
Wearn	.1069264	.0198877	5.38	0.000	.0679473	.1459055
_cons	-.0212474	.0047736	-4.45	0.000	-.0306035	-.0118913
micro						
ew	.6010344	.3251257	1.85	0.065	-.0362004	1.238269
vw	.3989656	.3251257	1.23	0.220	-.2382691	1.0362
variance						
lnsigma	-4.150034	.005244	-791.39	0.000	-4.160312	-4.139756

例

```
. mixregress ppi Lppi m0 if vw<=0.006, xmicro(earn) wmicro(ew vw) vce(2) idtype(1)
  nolog

Log-likelihood =       182.9834                   Number of obs    =     106454
AIC            =      -355.9667                   Number of group  =         67
BIC            =      -344.9433                   R-squared        =     0.8614
HQIC           =      -351.6047                   Adj R-squared    =     0.8501
DW             =         1.5601                   Root MSE         =     0.0165
```

ppi	Coefficient	HAC std. err.	z	P>|z|	[95% conf.	interval]
macro						
Lppi	.8923568	.0390872	22.83	0.000	.8157473	.9689663
m0	-.0003377	.000119	-2.84	0.005	-.0005709	-.0001044
Wearn	.0642664	.0408338	1.57	0.116	-.0157664	.1442993
_cons	-.0212474	.0047736	-4.45	0.000	-.0306035	-.0118913
micro						
ew	1	.	.	.	.	.
vw	.6637991	.9436116	0.70	0.482	-1.185646	2.513244
variance						
lnsigma	-4.150034	.005244	-791.39	0.000	-4.160312	-4.139756

例

```
. local r=1

. foreach v of varlist lnsale_w r_debtCF_w CashInEarn_w Tangibility_w {
  2. qui mixregress ppi Lppi if vw<=0.006, xmicro(earn) wmicro(ew vw `v') vce(2)
  3. est store mod`r'
  4. local r=`r'+1
  5. }
```

例

```
. est table mod1 mod2 mod3 mod4, star(.1 0.05 .01) stat(r2 r2a aic bic hqic)
```

Variable	mod1	mod2	mod3	mod4
macro				
Lppi	.89819419***	.87999128***	.91469816***	.90153225***
Wearn	.11262641***	.13102796***	.11936962***	.10771251***
_cons	-.02090985***	-.02072773***	-.0278286***	-.02118514***
micro				
ew	.56731637*	.5661791**	.56217911*	.59513198*
vw	.43268363	.4338209	.43782089	.40486802
lnsale_w	-1.593e-06			
r_debtCF_w		-.0008993**		
CashInEarn_w			9.178e-06	
Tangibilit~w				-3.332e-06
variance				
lnsigma	-4.1441738***	-4.1559908***	-4.1477835***	-4.1436328***
Statistics				
r2	.85981334	.86308766	.86082175	.85966156
r2a	.84832263	.85186533	.84941369	.84815841
aic	-353.18153	-354.76501	-353.66523	-353.10904
bic	-295.72872	-297.3122	-296.21242	-295.65623
hqic	-350.81952	-352.403	-351.30322	-350.74702

Legend: * p<.1; ** p<.05; *** p<.01

贝叶斯估计

```
. use "micro2.dta", clear
. qui merge m:1 yearquarter using macro2
. bysort yearquarter: gen iflast = _n==_N

. bayesmh ppi Lppi, ///
 llevaluator(lnfmixregress, parameters({th} {tau} {alpha} {var}) extravars(earn e
w vw Tangibility_w yearquarter iflast)) ///
 prior({ppi:Lppi}, beta(4,2))  prior({ppi:_cons}, normal(0, 10))  ///
 prior({th}, beta(4,2)) prior({tau}, normal(0,25))  ///
 prior({alpha}, normal(0,25))  prior({var}, igamma(0.01, 0.01))  ///
 initial({ppi:Lppi} 0.6 {ppi:_cons} 0 {th} 0.6 {tau} 0.5 {alpha} 0 {var} 100)  //
/
 block({var}) ///
 mcmcsize(2000) thinning(5) burnin(2500)
```

贝叶斯估计

```
Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  ppi ~ lnfmixregress(xb_ppi,{th},{tau},{alpha},{var})

Priors:
   {ppi:Lppi} ~ beta(4,2)                                                  (1)
  {ppi:_cons} ~ normal(0,10)                                               (1)
         {th} ~ beta(4,2)
        {tau} ~ exponential(10)
      {alpha} ~ normal(0,25)
        {var} ~ igamma(0.01,0.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_ppi.

Bayesian regression                              MCMC iterations  =      4,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =      2,000
                                                 Number of obs    =    107,425
                                                 Acceptance rate  =      .3151
                                                 Efficiency:  min =    .005055
                                                              avg =     .05403
```

```
Log marginal-likelihood =  125.09042                         max =       .2179
```

贝叶斯估计

```
------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
ppi          |
        Lppi |  .8907511   .0515945   .005436    .894502   .7761169    .967685
       _cons | -.0180184   .0050149   .000659  -.0174062  -.0282563  -.0086244
-------------+----------------------------------------------------------------
          th |  .6730623   .1842118   .027673   .6957836   .2869377   .9616901
         tau |  .2215724   .1485909   .046731   .1642277   .0691477   .5964011
       alpha |  .0007482    .000385    .00012    .000703   .0002096   .0015116
         var |  .0006607   .0001173   5.6e-06   .0006407   .0004638   .0009249
------------------------------------------------------------------------------
```

谢 谢！