小番茄

時間序列模型之價格預測分析

第三組
B06607058 邴國榮
B07607001 鄭鈞瀚
B07607015 楊皓閔
B07607027 詹勳亞

有無數量與建立模型上的RMSE及最適模型

- 無數量
- RMSE = 0.2682
- 最適模型：ARMA(2,1)

Number of obs	=	238
F(5, 233)	=	6834.77
Prob > F	=	0.0000
R-squared	=	0.9932
Adj R-squared	=	0.9931
Root MSE	=	.2682

- 有數量
- RMSE = 0.25614
- 最適模型：ARMA(2,1)

Number of obs	=	238
F(6, 232)	=	6248.66
Prob > F	=	0.0000
R-squared	=	0.9939
Adj R-squared	=	0.9937
Root MSE	=	.25614

Outline

- 背景介紹
- 資料分析
  - 月均資料
    - 分析過程（敘述統計、單根檢定、建立ARIMA模型、樣本內外預測、結論）

背景介紹

B

A

C

為什麼選小番茄

- 資料齊全，沒有missing value的存在
- 國內一年四季都有產
- 小番茄很好吃

圖片來源：台灣好水果 TWFOOD

資料分析

B

A

C

資料分析

- 農產品批發市場交易行情站（水果）
  - 產品交易價量走勢圖
- 日期別：月資料
  日期：2000.01 – 2019.12
- 價量：平均價與交易量
- 單位：公斤
- 市場：全部市場
- 產品：小番茄（一般）

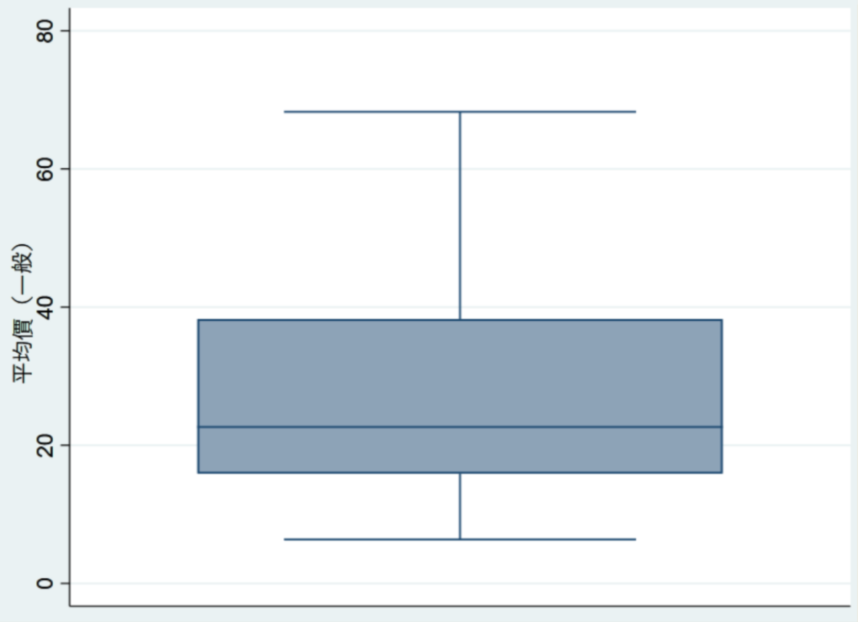
月均資料

敘述統計(1)

```
. summarize ave_price_1 log_P1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ave_price_1	240	27.32523	14.12244	6.358469	68.26905
log_P1	240	3.176793	.5171337	1.849788	4.223456

敘述統計(2)



box plot for “ave_price_1”

第一四分位數: 15.89299

中位數: 22.64355

第三四分位數: 38.27671

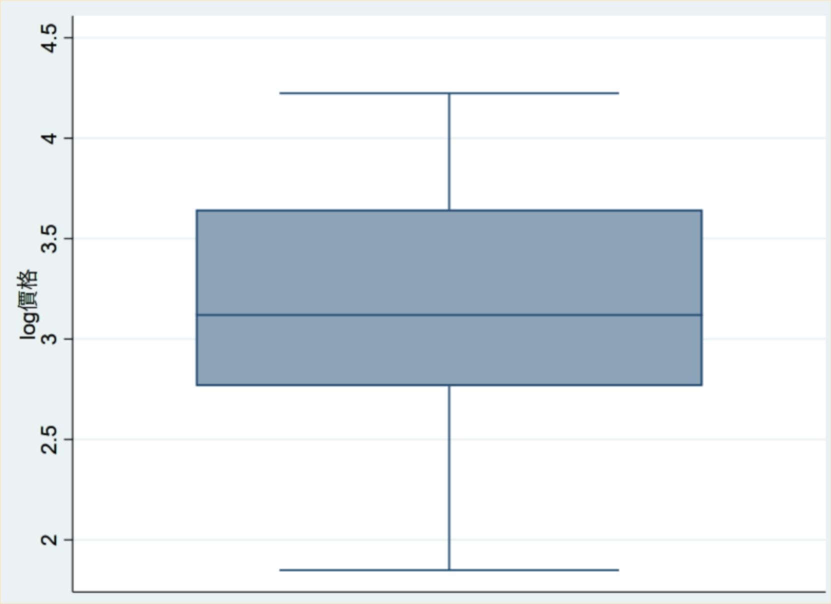
四分位距:22.38372

最小值: 6.358469

最大值: 68.26905

全距:61.910581

敘述統計(3)



box plot for “log_p”

第一四分位數:2.765868

中位數:3.119873

第三四分位數:3.644831

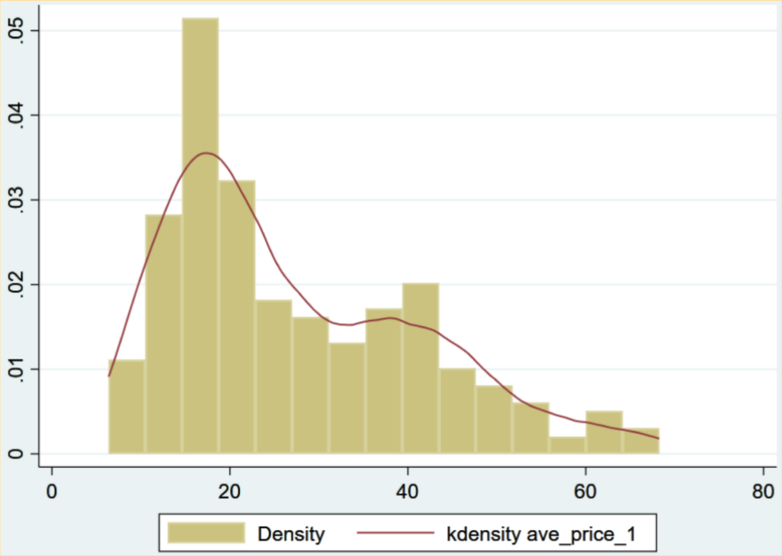
四分位距:0.878963

最小值:1.849788

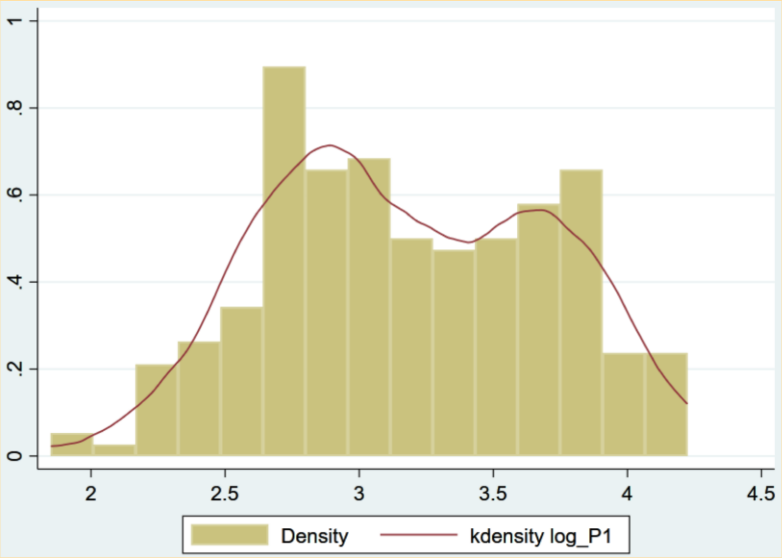
最大值: 4.223456

全距:2.373668

敘述統計(4)



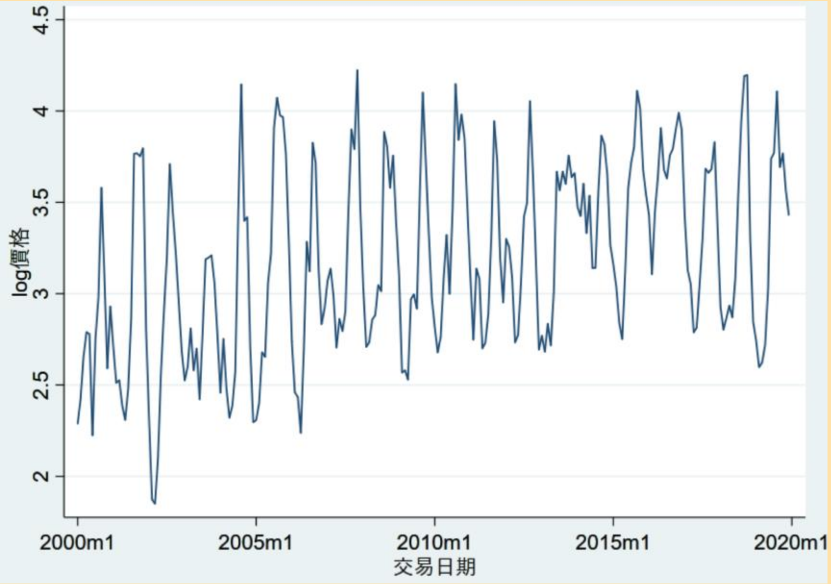
Histogram for “ave_price_1”



Histogram for “log_P1”

趨勢圖

- tsline log_P1 , name(unresi)
- 月均價格取對數後的趨勢圖



開始檢視資料是否為定態資料

- DICKEY-FULLER TEST
- PHILIPS-PERRON TEST
- KPSS TEST

單根檢定(1)
DICKY-FULLER TEST

- 檢定是否有單根
- H0 = 有單根 , H1 = 無單根
- Z(t) = -4.894 < 1% critical value
- Reject H0
- 此資料無單根

```
. do "C:\Users\user\AppData\Local\Temp\STD28a4_000000.tmp"

. dfuller log_P1 if date < tm(2017m1), regress

Dickey-Fuller test for unit root                   Number of obs   =       203

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -4.894            -3.476            -2.883            -2.573
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
    D.log_P1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      log_P1 |
         L1. |  -.2110241    .043123    -4.89   0.000    -.2960556   -.1259926
             |
       _cons |     .67266   .1376736     4.89   0.000     .4011901    .9441299
------------------------------------------------------------------------------
```

單根檢定(2)
DICKY-FULLER TEST

- 檢定是否有時間趨勢
  - H0：$\beta2 = 0$, H1：$\beta2 \neq 0$
  - $P > |t| = 0.011 < 0.05$
- 可看出有時間趨勢
- 檢定是否有shift term（常數項）
  - H0：$\beta0 = 0$, H1：$\beta0 \neq 0$
  - $P > |t| = 0.000 < 0.05$
- 可看出有shift term

Dickey-Fuller test for unit root Number of obs = 203

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-5.586	-4.006	-3.436	-3.136

MacKinnon approximate p-value for Z(t) = 0.0000

D.log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_P1						
L1.	-.2687852	.0481162	-5.59	0.000	-.3636654	-.1739049
_trend	.0010958	.0004267	2.57	0.011	.0002544	.0019373
_cons	.7428352	.1385191	5.36	0.000	.4696899	1.015981

單根檢定(3)
PHILIPS-PERRON TEST

- Z(rho)
  - H0 = 有單根, H1 = 無單根
  - Z(rho) = -70.741 < 1% critical value
  - Reject H0
  - 此資料無單根
- Z(t)
  - H0 = 有單根, H1 = 無單根
  - Z(rho) = -6.257 < 1% critical value
  - Reject H0
  - 此資料無單根

Phillips-Perron test for unit root Number of obs = 203
Newey-West lags = 4

	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
		Interpolated Dickey-Fuller		
Z(rho)	-70.741	-28.087	-21.112	-17.843
Z(t)	-6.257	-4.006	-3.436	-3.136

MacKinnon approximate p-value for Z(t) = 0.0000

log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
log_P1						
L1.	.7312148	.0481162	15.20	0.000	.6363346	.8260951
_trend	.0010958	.0004267	2.57	0.011	.0002544	.0019373
_cons	.7428352	.1385191	5.36	0.000	.4696899	1.015981

趨勢檢定

KPSS TEST

- 檢定是否有時間趨勢
  - Ho: Trend stationary,
    H1: Not trend stationary
- 0.052 < 10% critical value
- Do not reject H0
- The data is trend stationary

```
Maxlag = 8
Autocovariances weighted by Bartlett kernel

Critical values for H0: log_P1 is trend stationary

10%: 0.119  5% : 0.146  2.5%: 0.176  1% : 0.216

Lag order    Test statistic
    0            .0653
    1            .0377
    2            .0296
    3             .027
    4            .0271
    5            .0293
    6            .0338
    7            .0412
    8             .052
```

定勢檢定
KPSS TEST

- 檢定是否為定態
  - H0: level stationary
    H1: level unstationary
- 1.58 > 1% critical value
- Reject Ho
- 此資料為非定態

```
Maxlag = 8
Autocovariances weighted by Bartlett kernel

Critical values for H0: log_P1 is level stationary

10%: 0.347  5% : 0.463  2.5%: 0.574  1% : 0.739

Lag order    Test statistic
    0             4.42
    1             2.48
    2             1.87
    3             1.62
    4             1.51
    5             1.48
    6              1.5
    7             1.54
    8             1.58
```

單根檢定結論

	Dickey-Fuller	Philips-Perron	KPSS
單根	無	無	-
時間趨勢	有	有	有
是否定態	-	-	非定態

此資料無單根、有時間趨勢、非定態

Table 7.1 Restrictions on the overarching model

Zero-mean stationary AR(1):	$\beta_0 = 0$	$\|\beta_1\| < 1$	$\beta_2 = 0$
Non-zero mean stationary AR(1):	$\beta_0 \neq 0$	$\|\beta_1\| < 1$	$\beta_2 = 0$
Random Walk (RW):	$\beta_0 = 0$	$\beta_1 = 1$	$\beta_2 = 0$
Random Walk with Drift (RWWD):	$\beta_0 \neq 0$	$\beta_1 = 1$	$\beta_2 = 0$
Deterministic Trend (DT):		$\|\beta_1\| <= 1$	$\beta_2 \neq 0$

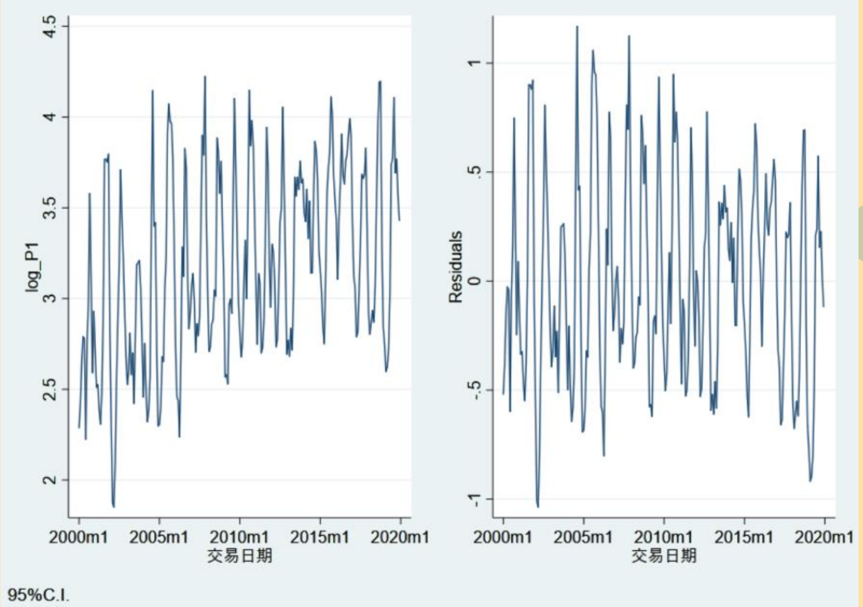
判斷此資料為
DETERMINISTIC TREND

DETREND

Notice that the first-differenced model now has an MA unit root in the error terms. Never take first differences to remove a deterministic trend. Rather, regress X on time, and then work with the residuals. These residuals now represent X that has been linearly detrended.

DETREND

- 利用殘差項將本資料轉為定態資料
- reg log_P1 date
- predict r_log_P1, residuals
- tsline r_log_P1, name(resi)
- graph combine unresi resi, note("95%C.I.")



Try and Error

DICKY-FULLER TEST

- H0=有單根,H1=無單根
- Z(t)= -5.586 < 1% critical value
- reject H0

- H0：β2 = 0 , H1：β2 ≠ 0
- P > |t| = 0.490 > 0.05
- 無時間趨勢

```
. dfuller r_log_P1 if date < tm(2017m1), regress trend
```

Dickey-Fuller test for unit root Number of obs = 203

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-5.586	-4.006	-3.436	-3.136

MacKinnon approximate p-value for Z(t) = 0.0000

D.r_log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
r_log_P1						
L1.	-.2687852	.0481162	-5.59	0.000	-.3636654	-.1739049
_trend	.0002632	.0003806	0.69	0.490	-.0004872	.0010137
_cons	-.0138124	.0445271	-0.31	0.757	-.1016152	.0739903

PHILIPS-PERRON TEST

- Z(rho), Z(t)
- 此資料無單根

- H0：β2 = 0 , H1：β2 ≠ 0
- P > |t| = 0.490 > 0.05
- 無時間趨勢

```
. pperron r_log_P1 if date < tm(2017m1), regress trend

Phillips-Perron test for unit root                 Number of obs   =       203
                                                   Newey-West lags =         4

                               ———————— Interpolated Dickey-Fuller ————————
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(rho)          -70.741           -28.087           -21.112           -17.843
 Z(t)             -6.257            -4.006            -3.436            -3.136
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
r_log_P1     |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    r_log_P1 |
         L1. |   .7312148   .0481162    15.20   0.000     .6363346    .8260951
      _trend |   .0002632   .0003806     0.69   0.490    -.0004872    .0010137
       _cons |  -.0138124   .0445271    -0.31   0.757    -.1016152    .0739903
------------------------------------------------------------------------------
```

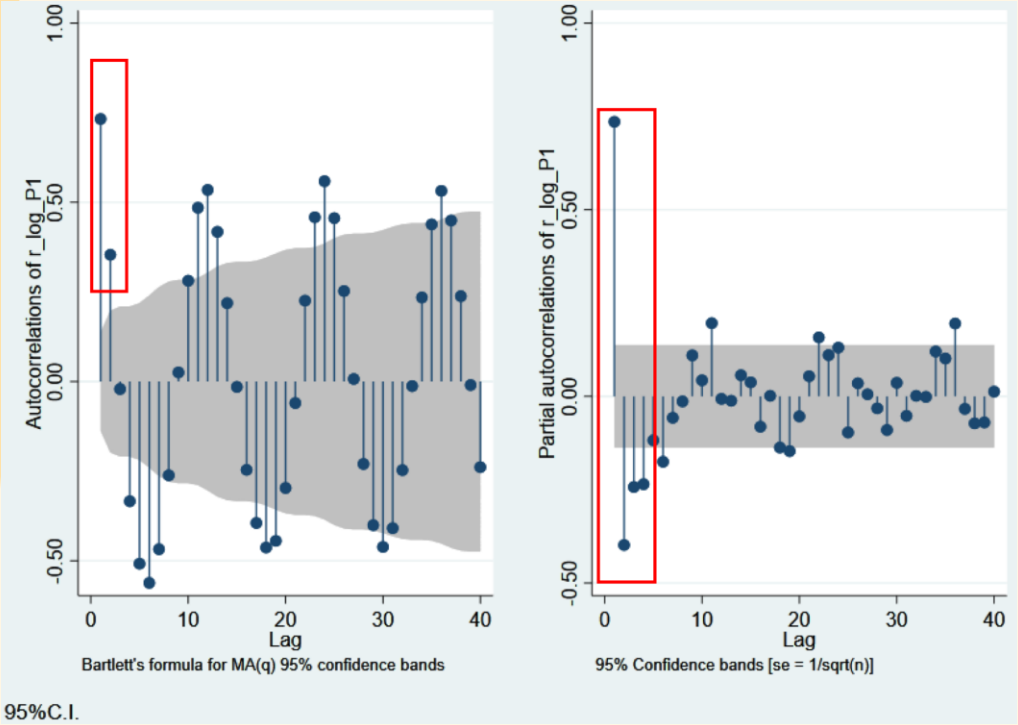
KPSS TEST

- 檢定是否為定態
  - H0: level stationary
    H1: level unstationary
  - 0.286 < 10% critical value
- Do not reject Ho
- 此資料為定態

```
. kpss r_log_P1 if date < tm(2017m1), maxlag(8) notrend

KPSS test for r_log_P1

Maxlag = 8
Autocovariances weighted by Bartlett kernel

Critical values for H0: r_log_P1 is level stationary

10%: 0.347  5% : 0.463  2.5%: 0.574  1% : 0.739

Lag order    Test statistic
    0             .389
    1             .225
    2             .176
    3             .159
    4             .159
    5             .171
    6             .195
    7             .232
    8             .286
```

ACF/PACF

- ac r_log_P1 if date <tm(2017m1), name(dac)
- pac r_log_P1 if date < tm(2017m1), name(dpac)
- graph combine dac dpac, note("95%C.I.")
- 結論：我們決定用ARMA(4,2)為基礎，建立arima模型



建立ARIMA模型(1)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(0,0,0) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-133.0337	1	268.0674	271.3855

```
Note: BIC uses N = number of observations. See [R] BIC note.

. qui arima r_log_P1 if date < tm(2017m1), arima(0,0,1) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-65.66813	2	135.3363	141.972

建立ARIMA模型(2)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(0,0,2) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        204          .  -47.90831       3    101.8166    111.771
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. qui arima r_log_P1 if date < tm(2017m1), arima(1,0,0) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        204          .  -53.19988       2    110.3998    117.036
-----------------------------------------------------------------------------
```

建立ARIMA模型(3)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(1,0,1) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-42.37674	3	90.75348	100.7078

Note: BIC uses N = number of observations. See [R] BIC note.

```
. qui arima r_log_P1 if date < tm(2017m1), arima(1,0,2) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-39.76132	4	87.52264	100.7951

建立ARIMA模型(4)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(2,0,0) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-35.79925	3	77.5985	87.55286

```
Note: BIC uses N = number of observations. See [R] BIC note.

. qui arima r_log_P1 if date < tm(2017m1), arima(2,0,1) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-22.52896	4	53.05792	66.3304

建立ARIMA模型(5)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(2,0,2) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-22.19279	5	54.38557	70.97617

Note: BIC uses N = number of observations. See [R] BIC note.

```
. qui arima r_log_P1 if date < tm(2017m1), arima(3,0,0) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-29.77257	4	67.54515	80.81763

建立ARIMA模型(6)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(3,0,1) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-22.30251	5	54.60502	71.19562

```
Note: BIC uses N = number of observations. See [R] BIC note.

. qui arima r_log_P1 if date < tm(2017m1), arima(3,0,2) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-21.79659	6	55.59319	75.50191

建立ARIMA模型(7)

```
. qui arima r_log_P1 if date < tm(2017m1), arima(4,0,0) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-24.21457	5	58.42914	75.01974

```
. qui arima r_log_P1 if date < tm(2017m1), arima(4,0,1) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-21.30933	6	54.61866	74.52738

```
. qui arima r_log_P1 if date < tm(2017m1), arima(4,0,2) nolog noconstant

. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	204	.	-20.68557	7	55.37113	78.59797

- 比較AIC, BIC最小者
- 決定使用ARIMA(2,0,1)

樣本內及樣本外預測

- 利用one_head、dynamic、t0進行預測

```
. predict one_head
(option xb assumed; predicted values)

. predict one_res, residual

. predict dynamic, dynamic(tm(2017m1))
(option xb assumed; predicted values)

. predict dynamic_res, dynamic(tm(2017m

. predict t0, t0(tm(2017m1))
(option xb assumed; predicted values)
(204 missing values generated)
```

樣本內及樣本外預測(1)

- 預測結果

實際值

估計值

```
. table period,contents(mean r_log_P1 mean one_head mean dynamic mean t0) format(%8.3f) row
```

period	mean(r_log_P1)	mean(one_head)	mean(dynamic)	mean(t0)
in sample	0.033	0.018	0.018	
out of sample	-0.185	-0.078	0.005	-0.066
Total	-0.000	0.003	0.016	-0.066

```
. table period,contents(sd r_log_P1 sd one_head sd dynamic sd t0) format(%8.3f)
```

period	sd(r_log_P1)	sd(one_head)	sd(dynamic)	sd(t0)
in sample	0.464	0.391	0.391	
out of sample	0.467	0.392	0.286	0.399

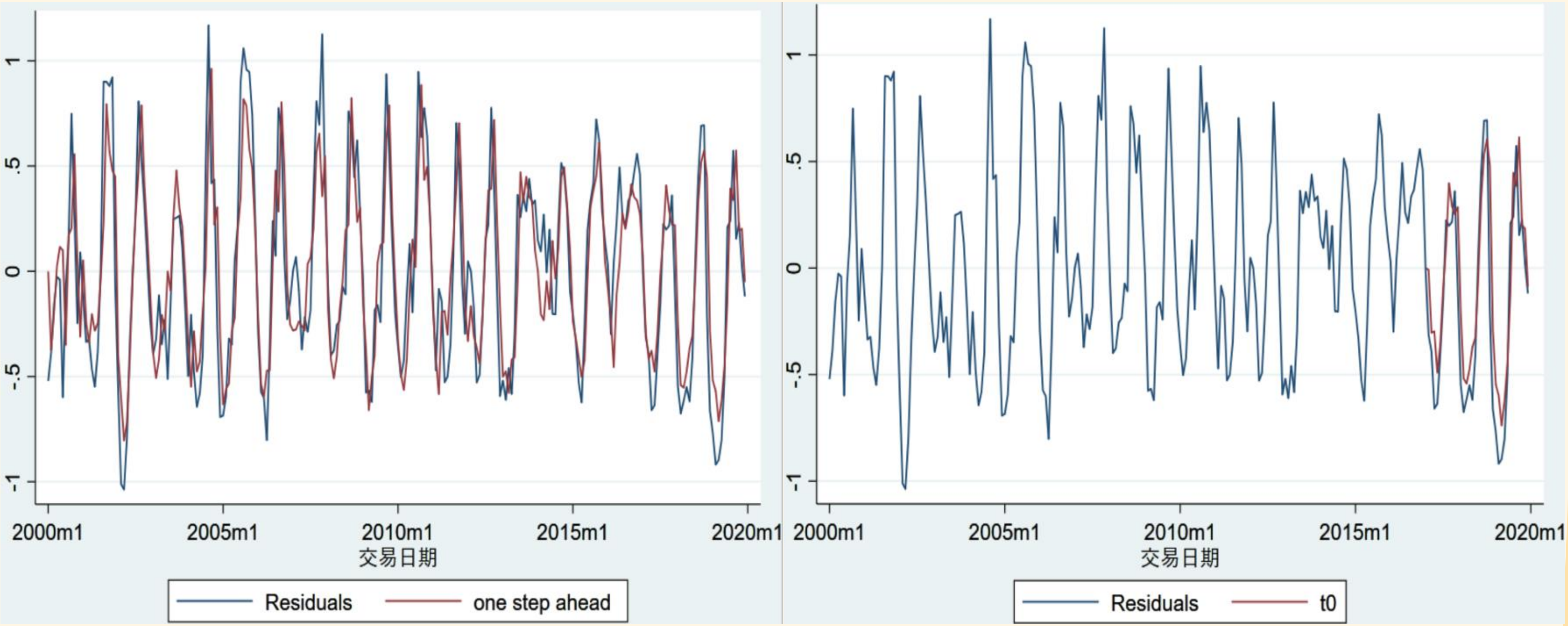
樣本內及樣本外預測(2)

```
. table period,contents(mean abs_one sd one_res mean abs_t0 sd  t0_res ) row format(%8.3f)
```

	MAE	MSE	MAE	MSE
period	mean(abs_one)	sd(one_res)	mean(abs_t0)	sd(t0_res)
in sample	0.202	0.257		
out of sample	0.182	0.201	0.188	0.207
Total	0.199	0.253	0.188	0.207

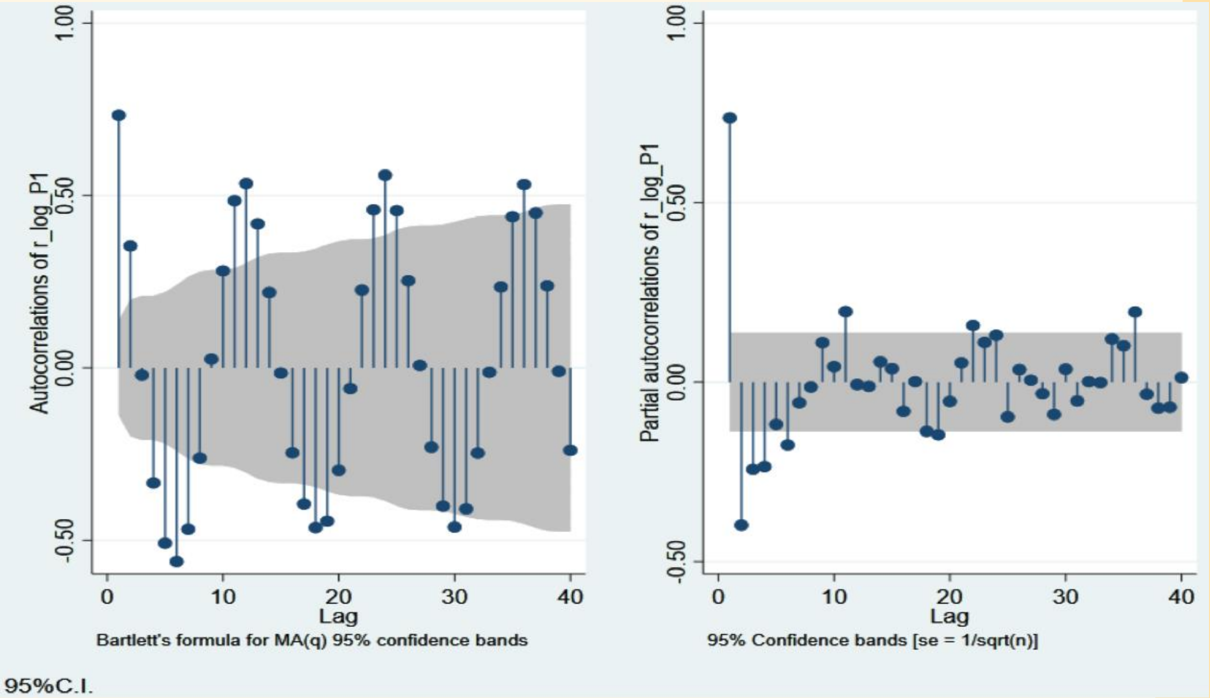
One-step-ahead　　　　t0

樣本內及樣本外預測(3)


1
.5
0
-.5
-1
2000m1
2005m1
2010m1
2015m1
2020m1
交易日期
Residuals
one step ahead
1
.5
0
-.5
-1
2000m1
2005m1
2010m1
2015m1
2020m1
交易日期
Residuals
t0

樣本內及樣本外預測(4)

- 資料可能偏誤因素
  - 資料在時間序列上存在週期性(Seasonality)
  - 未考慮價格的影響



ARIMA(2,0,1) 之方程式

- ARMA(2,1) with trend:
- $logP\_t=\beta\_1*logP\_(t-1)+\beta\_2*logP\_(t-2)+u\_t+\gamma\_1*u\_(t-1)+\alpha*t$

```
. twoway (tsline log_P1 ) (tsline xb2)

. /*Generate mse and mae*/
. reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1 date, noconstant
```

Source	SS	df	MS
Model	2458.14903	5	491.629806
Residual	16.7598588	233	.071930725
Total	2474.90889	238	10.3987768

Number of obs	=	238
F(5, 233)	=	6834.77
Prob > F	=	0.0000
R-squared	=	0.9932
Adj R-squared	=	0.9931
Root MSE	=	.2682

log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lag1_log_P1	10.8481	1.646151	6.59	0.000	7.60486	14.09134
lag2_log_P1	-9.846532	1.56515	-6.29	0.000	-12.93019	-6.762876
one_ahead	-11.53865	1.933207	-5.97	0.000	-15.34745	-7.729853
lag1_r_log_P1	8.515935	1.391011	6.12	0.000	5.775368	11.2565
date	-.0000538	.0006012	-0.09	0.929	-.0012383	.0011306

ARIMA(2,0,1) 之方程式

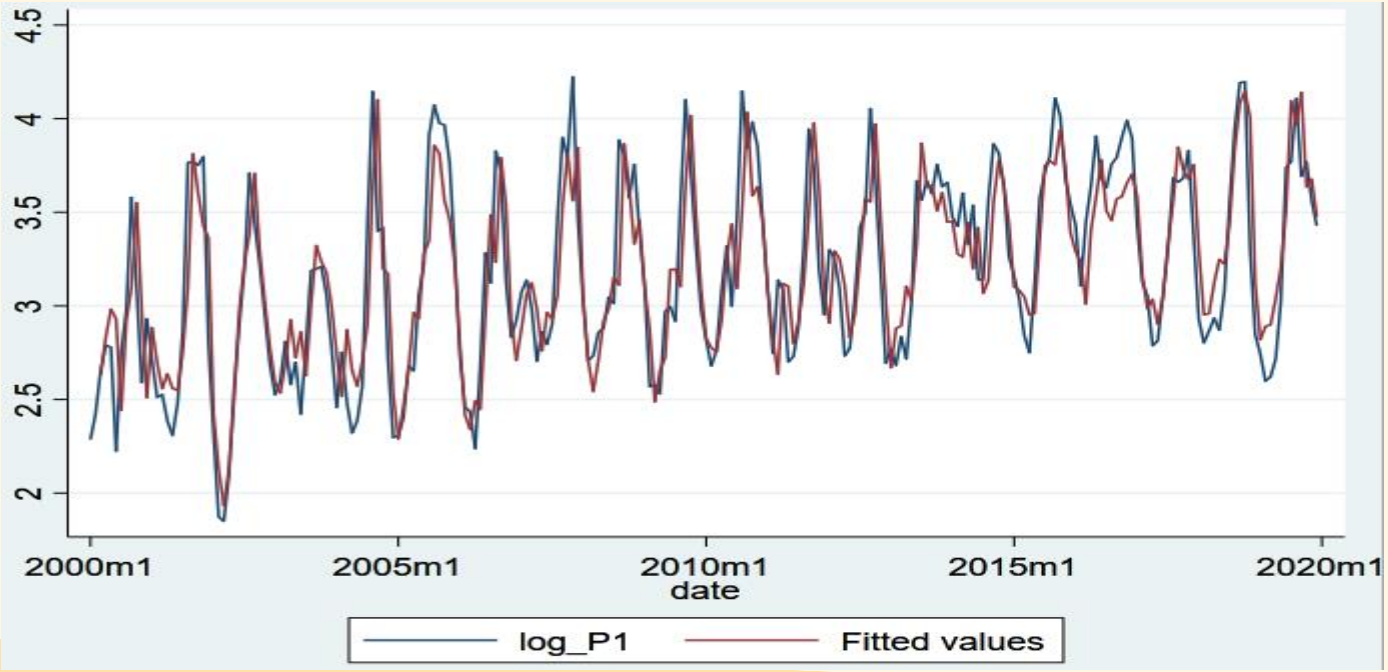
- ARMA(2,1) with trend:
  - logP_t=10.848*logP_(t-1)-9.847*logP_(t-2)-11.539u_t+8.516*u_(t-1)

```
. twoway (tsline log_P1 ) (tsline xb2)

. /*Generate mse and mae*/
. reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1 date, noconstant

      Source |       SS           df       MS      Number of obs   =       238
-------------+----------------------------------   F(5, 233)       =   6834.77
       Model |  2458.14903         5  491.629806   Prob > F        =    0.0000
    Residual |  16.7598588       233  .071930725   R-squared       =    0.9932
-------------+----------------------------------   Adj R-squared   =    0.9931
       Total |  2474.90889       238  10.3987768   Root MSE        =     .2682

------------------------------------------------------------------------------
      log_P1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lag1_log_P1 |    10.8481   1.646151     6.59   0.000      7.60486    14.09134
 lag2_log_P1 |  -9.846532    1.56515    -6.29   0.000    -12.93019   -6.762876
   one_ahead |  -11.53865   1.933207    -5.97   0.000    -15.34745   -7.729853
lag1_r_log_P1|   8.515935   1.391011     6.12   0.000     5.775368     11.2565
        date |  -.0000538   .0006012    -0.09   0.929    -.0012383    .0011306
------------------------------------------------------------------------------
```

- H0： α = 0，
  H1： α ≠ 0
- P > |t| = 0.929 > 0.05
- 無時間趨勢
- 與原先不符合

ARIMA(2,0,1) – 估計

- 對價格估計(one-step-ahead) 。



ARIMA(2,0,1) – MSE & MAE

- reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1, noconstant
- predict e,xb
- gen mse = (log_P1 -e)*(log_P1 -e)
- sum mse

```
. gen mse = ( log_P1 -e)*( log_P1 -e)
(2 missing values generated)

. sum mse
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mse	238	.0704196	.1010675	6.88e-09	.6003026

MSE(Mean square error)：0.0704196

- gen mae = abs(log_P1 -e)
- sum mae

```
. gen mae = abs( log_P1 -e)
(2 missing values generated)

. sum mae
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mae	238	.2081523	.1649439	.000083	.774792

MAE(Mean absolute error)：0.2081523

加入數量(1)

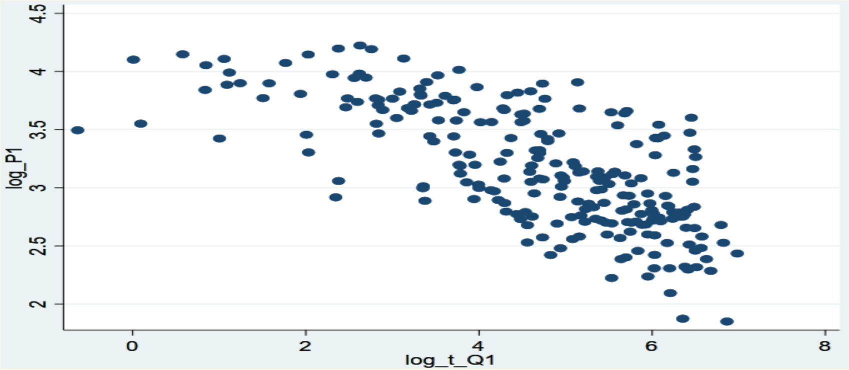
- gen tonnes_Q1 = quantity_1/1000
- gen log_t_Q1 = log(tonnes_Q1)
- summarize log_t_Q1

- 沒有missing value

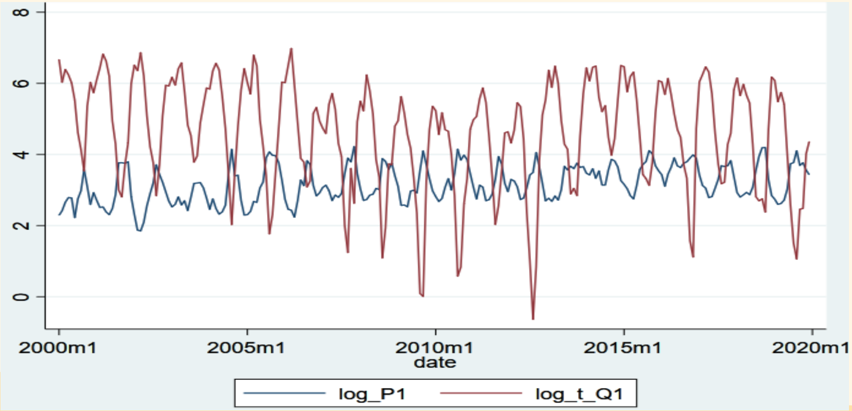
Variable	Obs	Mean	Std. Dev.	Min	Max
log_t_Q1	240	4.613669	1.534826	-.6348783	6.982008

加入數量(2)

- twoway scatter log_P1 log_t_Q1



- twoway (tsline log_P1) (tsline log_t_Q1)



- 由圖形推測兩者之間應該呈現線性關係

加入數量(3)

- 不過預測未來價格時我們不會知道外來的產量，因此選擇使用前一期來做判斷：

- reg log_P1 lag1_log_t_Q1, noconstant

- R-squared = 0.8087
  呈現高度相關

- H0：β = 0 , H1：β ≠ 0
- P > |t| = 0.000 < 0.05
- Reject H0
- 係數不等於0

```
. reg log_P1 lag1_log_t_Q1, noconstant
```

Source	SS	df	MS
Model	2006.13055	1	2006.13055
Residual	474.646505	238	1.99431305
Total	2480.77705	239	10.3798203

Number of obs	=	239
F(1, 238)	=	1005.93
Prob > F	=	0.0000
R-squared	=	0.8087
Adj R-squared	=	0.8079
Root MSE	=	1.4122

log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lag1_log_t_Q1	.5957411	.0187834	31.72	0.000	.5587381	.6327441

將前一期數量帶入並重新建立模型

- ARMA(2,1) with trend and quantity in previous period:
- logP_t=β_1*logP_(t-1)+β_2*logP_(t-2)+u_t+γ_1*u_(t-1)+α*t+ζ*log_Q(t-1)

```
. reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1 lag1_log_t_Q1 date, noconstant

      Source |       SS           df       MS      Number of obs   =       238
-------------+----------------------------------   F(6, 232)       =   6248.66
       Model |  2459.68836         6  409.94806   Prob > F        =    0.0000
    Residual |  15.2205287       232  .065605727   R-squared       =    0.9939
-------------+----------------------------------   Adj R-squared   =    0.9937
       Total |  2474.90889       238  10.3987768   Root MSE        =    .25614

------------------------------------------------------------------------------
      log_P1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 lag1_log_P1 |   15.24422   1.815267     8.40   0.000     11.66771    18.82074
 lag2_log_P1 |  -13.80528   1.703588    -8.10   0.000    -17.16176    -10.4488
   one_ahead |  -16.84664   2.146965    -7.85   0.000    -21.07668    -12.6166
lag1_r_log_P1|   11.88318   1.499336     7.93   0.000     8.929129    14.83724
lag1_log_t_Q1|  -.0847229   .0174906    -4.84   0.000    -.1191837   -.0502621
        date |  -.0017369   .0006711    -2.59   0.010    -.0030592   -.0004147
------------------------------------------------------------------------------
```

將前一期數量帶入並重新建立模型

- logP_t= 15.244*logP_(t-1)-13.805*logP_(t-2)-16.847*u_t+11.883*u_(t-1)-0.001*t-0.085ζ*log_Q(t-1)

```
. reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1 lag1_log_t_Q1 date, noconstant
```

Source	SS	df	MS
Model	2459.68836	6	409.94806
Residual	15.2205287	232	.065605727
Total	2474.90889	238	10.3987768

Number of obs	=	238
F(6, 232)	=	6248.66
Prob > F	=	0.0000
R-squared	=	0.9939
Adj R-squared	=	0.9937
Root MSE	=	.25614

log_P1	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lag1_log_P1	15.24422	1.815267	8.40	0.000	11.66771	18.82074
lag2_log_P1	-13.80528	1.703588	-8.10	0.000	-17.16176	-10.4488
one_ahead	-16.84664	2.146965	-7.85	0.000	-21.07668	-12.6166
lag1_r_log_P1	11.88318	1.499336	7.93	0.000	8.929129	14.83724
lag1_log_t_Q1	-.0847229	.0174906	-4.84	0.000	-.1191837	-.0502621
date	-.0017369	.0006711	-2.59	0.010	-.0030592	-.0004147

ARIMA(2,0,1) with Q – MSE & MAE

- reg log_P1 lag1_log_P1 lag2_log_P1 one_ahead lag1_r_log_P1 lag1_log_t_Q1 date, noconstant
- predict e_1,xb
- gen mse_Q1
  = (log_P1 -e_1)*(log_P1 -e_1)
- sum mse_Q1

- gen mae_Q1 = abs(log_P1 -e_1)
- sum mae_Q1

- 前一期數量的加入減少了誤差

```
. gen mse_Q1 = ( log_P1 -e_1)*( log_P1 -e_1)
(2 missing values generated)

. sum mse_Q1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mse_Q1	238	.0639518	.0936822	4.87e-07	.5249112

MSE(Mean square error)：0.0639518

```
. gen mae_Q1 = abs( log_P1 -e_1)
(2 missing values generated)

. sum mae_Q1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mae_Q1	238	.1985014	.1570114	.0006979	.7245076

MAE(Mean absolute error)：0.1985014

結論

- 單純樣本外預測，和現實偏差較大
- 預測的數值幅度較現實情況小
- 資料存在週期性(Seasonality)會影響到預測的判斷（ARMA不容易藉由AC與PAC得到）
- 數量的有無會影響到模型的準確：有數量比起無數量更準確

謝謝聆聽