Tesina 2 di modelli econometrici

Federico Spatola

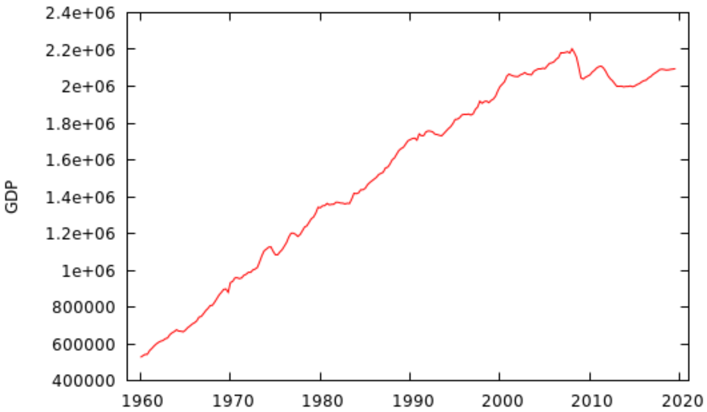
November 18, 2019

1 Consegna

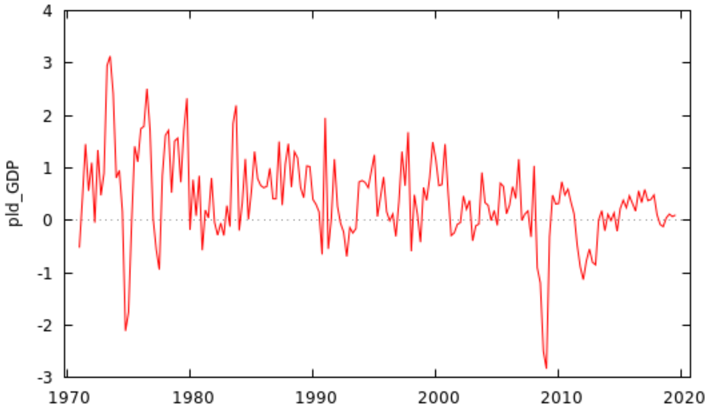
Scaricate i dati sul PIL trimestrale italiano a prezzi costanti e destagionalizzato (possibilmente gli ultimi dati disponiili su OECD http://stats.oecd.org) e tenendo conto di quanto appreso a lezione effettuate un'analisi secondo l'approccio di Box e Jenkins al fine di fare una previsione a uno, due, tre e quattro trimestri in avanti per il tasso di crescita trimestrale annualizzato del PIL in valori percentuali.

2 Analisi preliminare-Dataset 1960-2019

Si condurra' l'analisi su due differenti Dataset, per entrambi si faranno delle pseudo previsioni su uno stesso campione e si confronteranno i valori dell'RMFSE, infine utilizzando il modello con l'RMFSE minore si daranno le previsioni del tasso di crescita trimestrale annualizzato del PIL in valori percentuali. Adesso, si considera la serie storica del Pil ai prezzi costanti con dati destagionalizzati scaricabile da https://stats.oecd.org/Index.aspx?DataSetCode=QNA#.

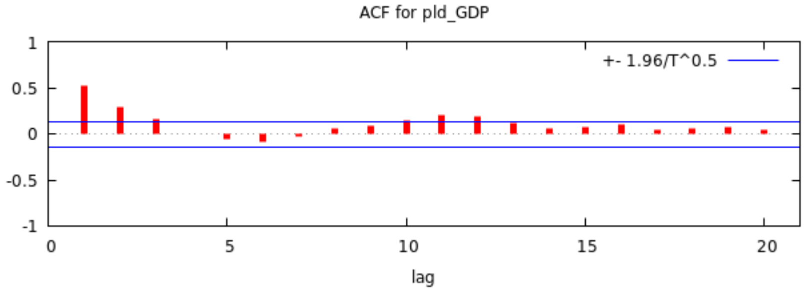


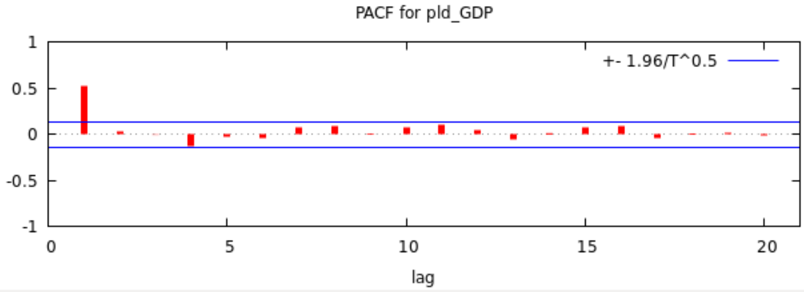
Si nota la crescita logaritmica e quindi una forte asimmetria, per ridurla si passa alla serie dei logaritmi poi alla serie delle differenze prime di quest'ultima e infine si moltiplica per 100 , ottenendo un approssimazione del tasso di crescita trimestrale del PIL in valori percentuali, da notare che moltiplicando ancora per 4 si potrebbe effettuare l'analisi direttamente sul tasso trimestrale annualizzato. Tenendo l'intero campione Il Test di Burtlett mostra che c'e' una forte correlazione tra i ritardi, allora si e' deciso di eliminare le osservazioni fino al primo trimestre del 1971, cosi' facendo si ottiene un considerevole miglioramento nella stazionarieta' della serie:



Di seguito il correlogramma :

```
Autocorrelation function for pld_GDP
***, **, * indicate significance at the 1%, 5%, 10% levels
using standard error 1/T^0.5

  LAG      ACF          PACF         Q-stat. [p-value]

    1   0.5236  ***   0.5236 ***     54.2925  [0.000]
    2   0.3000  ***   0.0355         72.2004  [0.000]
    3   0.1674  **   -0.0038         77.8040  [0.000]
    4   0.0002       -0.1291 *       77.8040  [0.000]
    5  -0.0565       -0.0203         78.4493  [0.000]
    6  -0.0878       -0.0331         80.0171  [0.000]
    7  -0.0243        0.0843         80.1381  [0.000]
    8   0.0683        0.0962         81.0977  [0.000]
    9   0.0878        0.0107         82.6904  [0.000]
   10   0.1518  **    0.0824         87.4726  [0.000]
   11   0.2096  ***   0.1005         96.6421  [0.000]
   12   0.1975  ***   0.0438        104.8294  [0.000]
   13   0.1157       -0.0480        107.6538  [0.000]
   14   0.0699        0.0148        108.6898  [0.000]
   15   0.0805        0.0788        110.0737  [0.000]
   16   0.1020        0.0901        112.3045  [0.000]
   17   0.0539       -0.0319        112.9310  [0.000]
   18   0.0564        0.0121        113.6201  [0.000]
   19   0.0757        0.0202        114.8695  [0.000]
   20   0.0478       -0.0173        115.3708  [0.000]
```





3 Identificazione

Il correlogramma suggerisce che la serie storica si possa identificare con un modello AR(ρ). Si sceglie $\rho = 12$, si considerano i modelli AR(0),..,AR(12) e dunque si utilizzano i criteri informativi per verificare quale di questi si adatta meglio ai dati, in particolare si usa in Gretl la funzionalita' Var Lag Selection.

```
VAR system, maximum lag order 12

The asterisks below indicate the best (that is, minimized) values
of the respective information criteria, AIC = Akaike criterion,
BIC = Schwarz Bayesian criterion and HQC = Hannan-Quinn criterion.

lags        loglik    p(LR)       AIC          BIC          HQC

   1    -190.22235             2.100791*    2.135867*    2.115009*
   2    -190.15597  0.71560    2.110994     2.163609     2.132321
   3    -190.15524  0.96946    2.121915     2.192068     2.150352
   4    -188.22018  0.04915    2.111696     2.199387     2.147241
   5    -188.09356  0.61481    2.121241     2.226470     2.163896
   6    -187.97804  0.63075    2.130908     2.253675     2.180671
   7    -187.68839  0.44659    2.138671     2.278976     2.195544
   8    -186.58066  0.13663    2.137494     2.295337     2.201475
   9    -186.43464  0.58892    2.146827     2.322208     2.217917
  10    -185.41984  0.15426    2.146665     2.339585     2.224865
  11    -184.31618  0.13736    2.145532     2.355990     2.230841
  12    -184.08023  0.49211    2.153882     2.381879     2.246300
```

Entrambi i criteri AIC e BIC suggeriscono di scegliere AR(1).

4 modello AR(1)

Di seguito l'output OLS relativo al modello AR(1).

```
Model 4: OLS, using observations 1971:2-2019:3 (T = 194)
Dependent variable: pld_GDP

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       0.194602      0.0568534     3.423    0.0008   ***
  pld_GDP_1   0.523964      0.0612289     8.557    3.65e-15 ***

Mean dependent var   0.405255   S.D. dependent var   0.836778
Sum squared resid    97.82645   S.E. of regression   0.713802
R-squared            0.276101   Adjusted R-squared   0.272331
F(1, 192)            73.23030   P-value(F)           3.65e-15
Log-likelihood      -208.8617   Akaike criterion     421.7235
Schwarz criterion    428.2592   Hannan-Quinn         424.3700
rho                 -0.014149   Durbin's h          -0.377388
```

5 Diagnostiche

5.1 Test di autocorrelazione

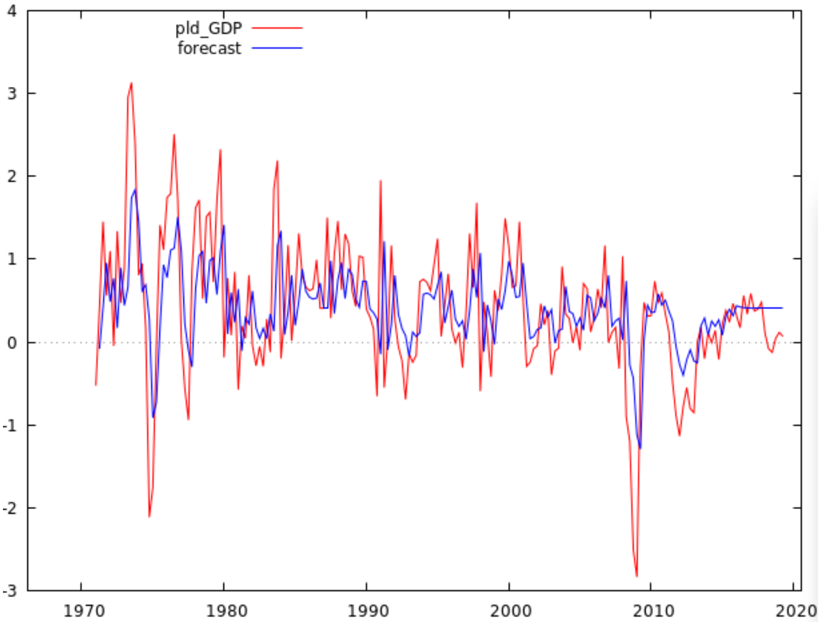
Di seguito, si allega l'LMTest di autocorrelazione.

```
LM test for autocorrelation up to order 1 -
  Null hypothesis: no autocorrelation
  Test statistic: LMF = 0.136319
  with p-value = P(F(1, 191) > 0.136319) = 0.712377

LM test for autocorrelation up to order 5 -
  Null hypothesis: no autocorrelation
  Test statistic: LMF = 0.987305
  with p-value = P(F(5, 187) > 0.987305) = 0.426792
```

Si accetta l' ipotesi di non autocorrelazione.

6 Previsione

Si comincia effettuando una pseudo-previsione, lasciando fuori dal campione di stima le osservazioni successive al primo trimestre del 2016.



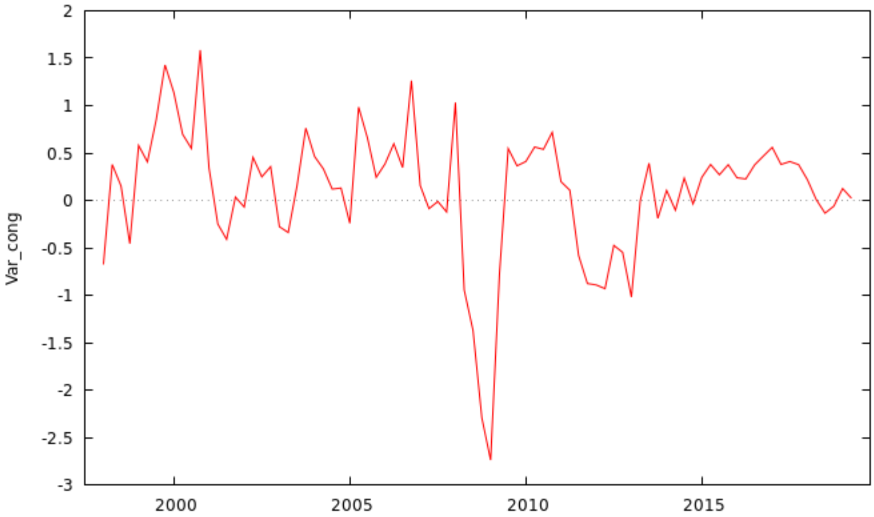

```
For 68% confidence intervals, t(192, 0.16) = 0.997

             pld_GDP    prediction    std. error        68% interval

2016:1      0.320004     0.438739
2016:2      0.173952     0.424486
2016:3      0.562286     0.417018
2016:4      0.337373     0.413105
2017:1      0.586971     0.411055
2017:2      0.375569     0.409980
2017:3      0.397395     0.409417
2017:4      0.481712     0.409123
2018:1      0.105888     0.408968
2018:2     -0.079237     0.408887
2018:3     -0.122701     0.408845
2018:4      0.046755     0.408822
2019:1      0.117247     0.408811
2019:2      0.069557     0.408805

  Forecast evaluation statistics

  Mean Error                       -0.17238
  Root Mean Squared Error           0.2784
  Mean Absolute Error               0.22863
  Mean Percentage Error            -63.762
  Mean Absolute Percentage Error    223.78
  Theil's U                         2.9385
  Bias proportion, UM               0.38337
  Regression proportion, UR         0.019239
  Disturbance proportion, UD        0.59739
```

Risulta che l' RMFSE e' uguale a 0.2784.

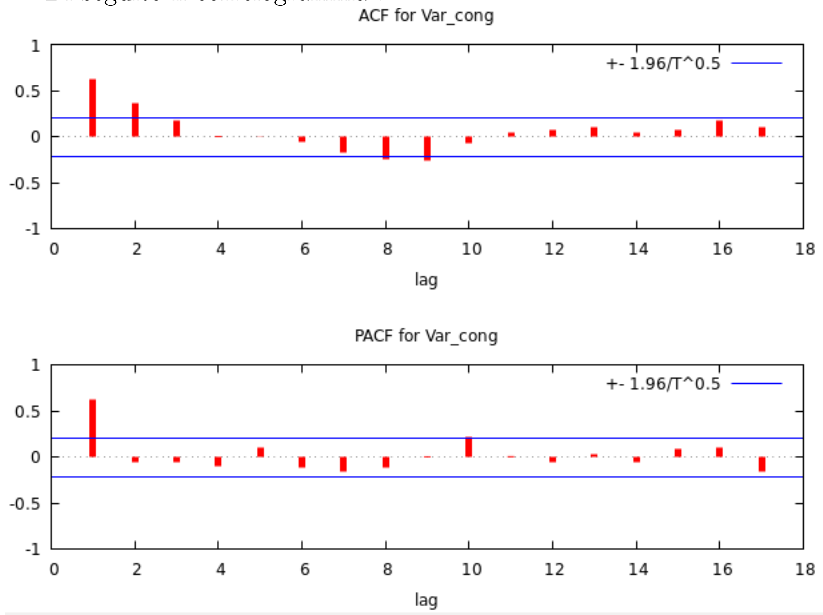
7 Analisi preliminare

Adesso si considera la serie storica della variazione percentuale congiunturale del Pil ai prezzi di mercato con dati destagionalizzati, scaricabile da http://dati.istat.it/.

$$\text{Variazione percentuale} = \left(\frac{x_f - x_i}{x_i} \cdot 100\right)\%$$



Di seguito il correlogramma :



Il Test di Burtlett rifiuta al 5 per cento, l' ipotesi che l'autocorrelazione sia nulla per k=1,2,8,9.

8 Identificazione

Si procede analogamente a come si e' fatto per il primo Dataset, si suppone che il modello vero sia un AR(ρ).Si sceglie $\rho = 9$, si considerano i modelli AR(0),..,AR(9) e dunque si utilizzano i criteri informativi per verificare quale di questi si adatta meglio ai dati. La prima operazione da effettuare e' eliminare le prime 9 osservazioni. Dopo di che si stimano i modelli OLS con differente numero di ritardi.

AR(0) Akaike criterion 157.5466 Schwarz criterion 159.8904
AR(1) Akaike criterion 118.5814 Schwarz criterion 123.2690
AR(2) Akaike criterion 119.8479 Schwarz criterion 126.8793
AR(3) Akaike criterion 121.3995 Schwarz criterion 130.7747
AR(4) Akaike criterion 123.3005 Schwarz criterion 135.0195
AR(5) Akaike criterion 124.8252 Schwarz criterion 138.8880
AR(6) Akaike criterion 125.7703 Schwarz criterion 142.1770
AR(7) Akaike criterion 125.8305 Schwarz criterion 144.5809
AR(8) Akaike criterion 126.8732 Schwarz criterion 147.9674
AR(9) Akaike criterion 128.8724 Schwarz criterion 152.3104 .

Si sceglie ancora il modelloAR(1).

9 modello AR(1)

Di seguito l'output OLS relativo al modello AR(1).

```
Model 23: OLS, using observations 1998:2-2019:2 (T = 85)
Dependent variable: Var_cong

              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        0.0463612    0.0574077    0.8076    0.4216
  Var_cong_1   0.631583     0.0839892    7.520     5.83e-11 ***

Mean dependent var   0.111741   S.D. dependent var   0.674314
Sum squared resid    22.71748   S.E. of regression   0.523168
R-squared            0.405220   Adjusted R-squared   0.398054
F(1, 83)             56.54747   P-value(F)           5.83e-11
Log-likelihood      -64.53032   Akaike criterion     133.0606
Schwarz criterion    137.9459   Hannan-Quinn         135.0256
rho                  0.063269   Durbin's h           0.921849
```

10 Diagnostiche

10.1 Test di autocorrelazione

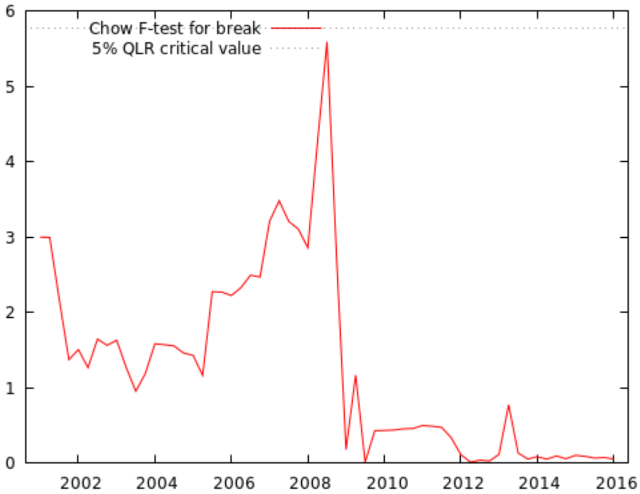
Di seguito, si allega l'LMTest di autocorrelazione.

```
LM test for autocorrelation up to order 5 -
  Null hypothesis: no autocorrelation
  Test statistic: LMF = 0.755854
  with p-value = P(F(5, 78) > 0.755854) = 0.584352

LM test for autocorrelation up to order 1 -
  Null hypothesis: no autocorrelation
  Test statistic: LMF = 0.79898
  with p-value = P(F(1, 82) > 0.79898) = 0.374014
```

Si accetta l' ipotesi di non autocorrelazione.

10.2 QLR-Test

Si esegue il QLR-Test.

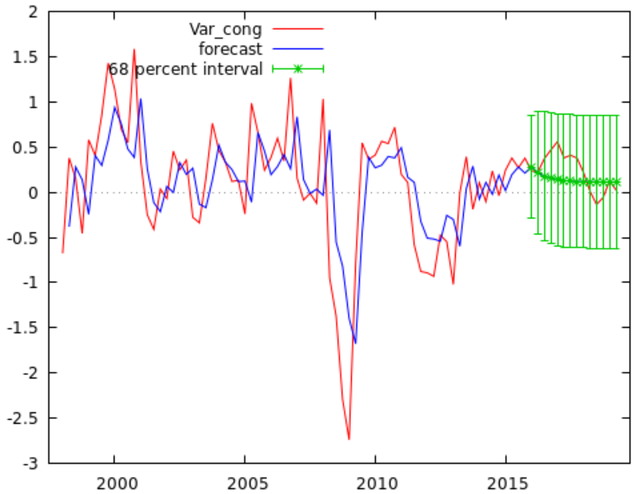



```
Quandt likelihood ratio test for structural break at an unknown point,
with 15 percent trimming:
The maximum F(2, 81) = 5.58982 occurs at observation 2008:3
Asymptotic p-value = 0.0594151 for chi-square(2) = 11.1796
```

Il Test ci fornisce il trimestre in cui si verifica il break strutturale. Se ne terra' conto in seguito.

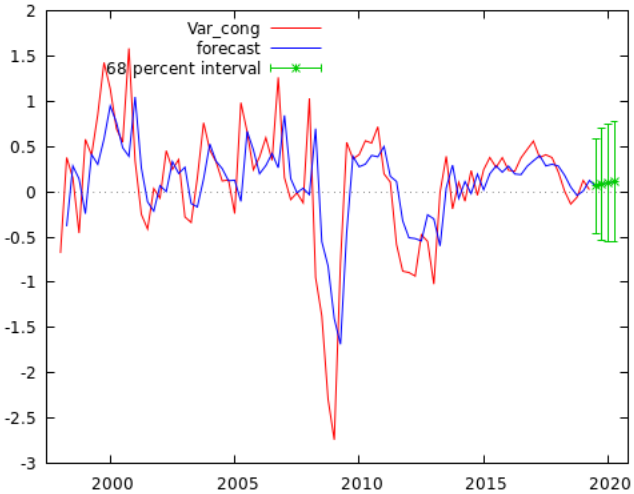
11 Previsione

Si comincia effettuando una pseudo-previsione, lasciando fuori dal campione di stima le osservazioni successive al primo trimestre del 2016 quindi circa il 20 per cento del campione.




```
For 68% confidence intervals, t(69, 0.16) = 1.002

          Var_cong    prediction    std. error        68% interval

2016:1       0.239         0.279        0.5708      -0.293 -     0.850
2016:2       0.227         0.217        0.6738      -0.458 -     0.892
2016:3       0.375         0.178        0.7102      -0.533 -     0.890
2016:4       0.469         0.154        0.7241      -0.571 -     0.879
2017:1       0.560         0.139        0.7294      -0.592 -     0.870
2017:2       0.380         0.129        0.7315      -0.603 -     0.862
2017:3       0.410         0.123        0.7324      -0.610 -     0.857
2017:4       0.378         0.120        0.7327      -0.614 -     0.854
2018:1       0.219         0.117        0.7328      -0.617 -     0.851
2018:2       0.010         0.116        0.7329      -0.618 -     0.850
2018:3      -0.134         0.115        0.7329      -0.619 -     0.849
2018:4      -0.060         0.114        0.7329      -0.620 -     0.848
2019:1       0.126         0.114        0.7329      -0.620 -     0.848
2019:2       0.024         0.114        0.7329      -0.620 -     0.848

  Forecast evaluation statistics

  Mean Error                           0.085252
  Root Mean Squared Error              0.21556
  Mean Absolute Error                  0.17931
  Mean Percentage Error               -36.626
  Mean Absolute Percentage Error       170.28
  Theil's U                            1.6873
  Bias proportion, UM                  0.15641
  Regression proportion, UR            7.3265e-05
  Disturbance proportion, UD           0.84352
```

L'RMSE e' piu' piccolo di quello trovato sul Dataset precedente, per questo motivo si sceglie di basare le previsioni, richieste nella consegna, su questo Dataset. Usando tutte le osservazioni si da' la seguente previsione dinamica quattro trimestri in avanti:



For 68% confidence intervals, t(83, 0.16) = 1.000

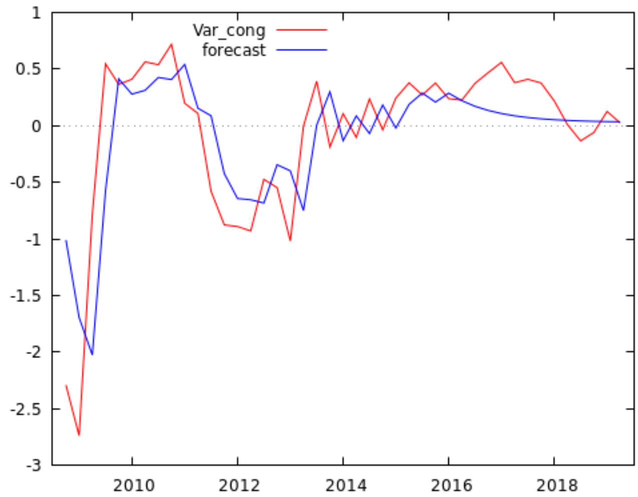
	Var_cong	prediction	std. error	68% interval
2019:3		0.062	0.5232	-0.462 - 0.585
2019:4		0.085	0.6188	-0.534 - 0.704
2020:1		0.100	0.6530	-0.553 - 0.753
2020:2		0.110	0.6662	-0.557 - 0.776

Le previsioni corrispondono alle variazioni percentuali congiunturali previste dal modello. Se si vuole ottenere la stima del PIL nel terzo trimestre del 2019 si considera il valore nel secondo trimestre (secondo i dati del dataset..) e si calcola:

$$2093509 + \frac{2093509 \cdot \text{prediction}}{100} = 2093509 + \frac{2093509 \cdot 0.062}{100} = 2094953.52.$$

Analogamente si stimano i valori del Pil previsti nei successivi tre trimestri.

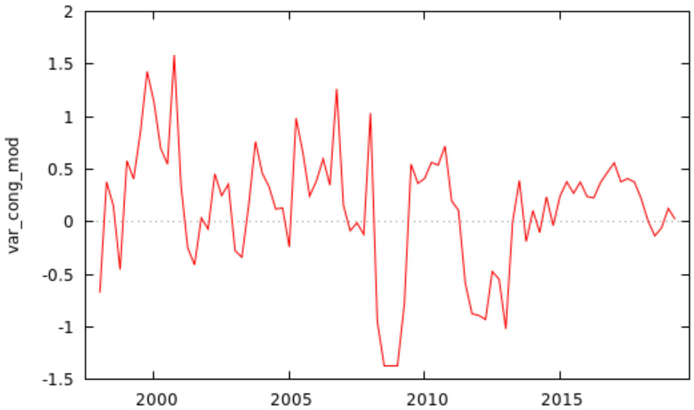
Adesso, si sceglie di considerare solo il campione successivo al break strutturale.



Forecast evaluation statistics

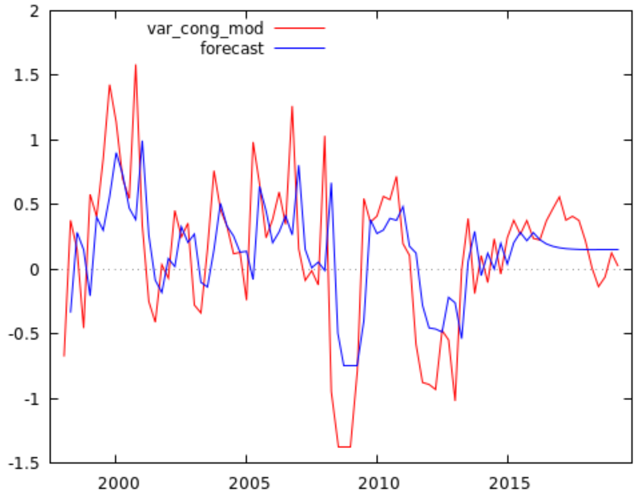
Mean Error	0.13073
Root Mean Squared Error	0.22967
Mean Absolute Error	0.18357
Mean Percentage Error	33.314
Mean Absolute Percentage Error	94.595
Theil's U	1.2015
Bias proportion, UM	0.324
Regression proportion, UR	6.6759e-05
Disturbance proportion, UD	0.67593

Si nota che l'RMSE e' leggermente peggiorato, probabilmente poiche' il campione di stima e' risultato troppo piccolo. Si prova adesso un altra strada ovvero si prova ad eliminare le osservazioni outliers 2008:4 e 2009:1, supponendo quindi che il trend osservato in quei trimestri non si verifichera' nei prossimi anni. In particolare si suppone che i due trimestri abbiano registrato la medesima crescita del trimestre 2008:3. Si ottiene la serie storica in figura.

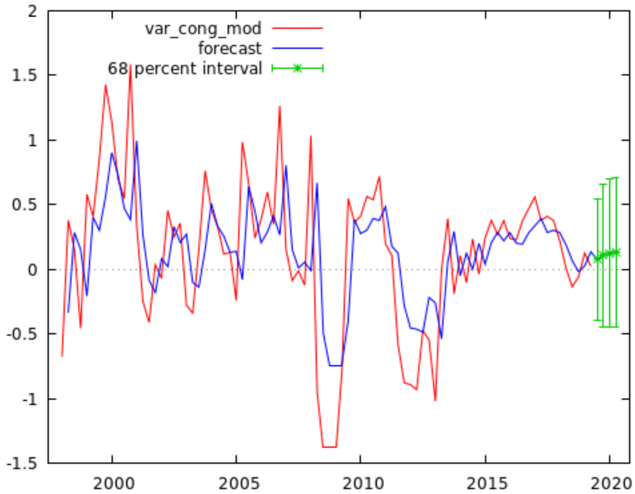


Utilizzando sempre il modello AR(1) si ottengono le pseudo-previsioni:

```
Forecast evaluation statistics

Mean Error                       0.056682
Root Mean Squared Error          0.20769
Mean Absolute Error              0.176
Mean Percentage Error           -73.432
Mean Absolute Percentage Error   209.69
Theil's U                        1.9317
Bias proportion, UM              0.074484
Regression proportion, UR        0.00028986
Disturbance proportion, UD       0.92523
```



L'RMSE di questo modello e' il risultato migliore ottenuto. Si da' una nuova previsione dinamica quattro trimestri in avanti.




```
For 68% confidence intervals, t(83, 0.16) = 1.000

       var_cong_mod    prediction    std. error        68% interval

2019:3                      0.076        0.4721      -0.396 -     0.548
2019:4                      0.107        0.5477      -0.441 -     0.655
2020:1                      0.125        0.5716      -0.447 -     0.697
2020:2                      0.135        0.5796      -0.445 -     0.715
```

Dunque le stime che si danno per i tassi di crescita trimestrali annualizzati in termini percentuali per i prossimi 4 trimestri sono rispettivamente $4 \cdot 0.076\% = 0.304\%, 0.428\%, 0.5\%$ e 0.540%.