Paper 204

Testing for Heteroscedasticity in the Multivariate ENSO Index Using PROC AUTOREG

Joseph Earley, Loyola Marymount University, Los Angeles

Abstract

El Nino/Southern Oscillation [ENSO] is a disruption in the ocean-atmosphere system in the tropical Pacific having important consequences for weather around the globe.[1] The purpose of this paper is to illustrate how the SAS® System may be used to gain insight into the nature of the Multivariate ENSO Index (MEI) which is used in modeling and forecasting ENSO activity. The paper presents the results of using the SAS PROC AUTOREG procedure to test for heteroscedasticity in the MEI time series. The statistical problem of heteroscedasticity, i.e. unequal error variance, is a violation of one of the assumptions upon which the least-squares method of regression is based. Heteroscedasticity must be taken into account in the estimation process for estimates to be efficient and not over-parameterized. Using the results of the Portmanteau Q-statistic and the Engle-Lagrange multiplier (LM) tests, the MEI series is then estimated using the family of generalized autoregressive conditional heteroscedasticity (GARCH) models. In addition to the statistical analysis using SAS/ETS procedures, a brief discussion of the El Nino phenomenon is presented.

What is El Nino?

The term El Nino (which is Spanish for the Christ Child) was originally used by fishermen along the coasts of Ecuador and Peru to refer to a warm ocean current that typically appears around Christmas time and lasts for several months duration.[2] The term El Nino has come to mean exceptionally strong warm intervals that bring heavy rains. There have been nine El Ninos in the past 40 years - the strong ones such as the 1982-83 El Nino effecting both weather conditions and marine life around the globe. The relationship between El Nino activity and resulting climatic effects has been well established. During the 1920's Sir Gilbert Walker, a British scientist, discovered a relationship between barometer readings at stations on the eastern and western sides of the Pacific. In particular, Walker noticed that when pressure rises in the east, it usually falls in the west, and vice versa. This phenomenon was called the Southern Oscillation. In the late 1960's, Jacob Bjerknes, a University of California professor noted a relationship between unusually warm sea-surface temperatures and the weak easterlies and heavy rainfall that accompany low-index conditions. Bjerknes' discovery led to the recognition that the warm waters of El Nino and the Walker barometer changes are part of the same phenomenon which today is called the El Nino - Southern Oscillation, or ENSO.

Multivariate ENSO Index (MEI)

In order to capture the relationship between El Nino and the Southern Oscillation, scientists at the NOAA-CIRES Climate Diagnostics Center(CDC), at the University of Colorado at Boulder, calculate a multivariate ENSO Index (MEI), which uses six main observable variables over the tropical Pacific. These six variables are: sea-level pressure (P), zonal (U) and meridional (V) components of the surface wind, sea surface temperature(S), surface air temperature(A), and total cloudiness fraction of the sky(C). The MEI is then computed separately for each of twelve sliding bi-monthly seasons. It is calculated as the first unrotated principal component of all six observed fields combined.[3]

[1] See: http://www.pmel.noaa.gov/toga-tao/el-ninostory.html

[2] This section draws from the website: http://www.pmel.noaa.gov/toga-tao/el-nino-report.html, a publication of the National Oceanic and Atmospheric Administration(NOAA).

[3] See Wolter, K., and Timlin, M.S, 1993 Monitoring ENSO in COADS with a seasonally adjusted principal component index. Proceedings of the 17th Climate Diagnostics Workshop, Normal, OK, NOAA/N MC/CAC, NSSL, Oklahoma Climate Survey, CIMMS and the School of Meteorology, University of Oklahoma, 52-57.

ARCH/GARCH Modeling[4]

The first researcher to develop a method for estimating time series with heteroscedasticity was R.F. Engle in 1982. Since then, numerous researchers such as Bollerslev (1986) have refined the method and today it is called GARCH - generalized autoregressive conditional heteroscedasticity modeling. Unlike the ARIMA model which assumes that the time series has (or has been transformed into) a constant variance, the ARCH/GARCH model does not make that assumption. Instead, the ARCH/GARCH model uses past squared residuals to estimate the current error variance in its modeling and forecasting. The ARCH model is used for modeling short memory processes while the GARCH model is used for long memory processes. The paper will illustrate how to determine the type and order of the appropriate ARCH/GARCH process.

Heteroscedasticity and ARCH/GARCH Modeling with PROC AUTOREG

An ARCH(q) model may be expressed as:

$$Y_t = \varepsilon_t \sqrt{h_t}$$

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \ \varepsilon^2_{t-i}$$

where $y_t \mid \psi_{t-1} \sim N(0, h_t)$ and

ψ_{t-1} is the information set at time t-1.

The GARCH (p,q) model where the conditional variance is expressed in terms of both lagged sample variances as well as past values of the conditional variance may be expressed as:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \ \varepsilon^2_{t-i} + \sum_{j=1}^{p} \gamma_j \ h_{t-j}$$

where: $p \geq 0$, $q>0$, $\omega>0$, $\alpha_i \geq 0$, $\gamma_j \geq 0$.

[4] See Woodward, Donna (1995) An Introduction to ARCH/GARCH Modeling Using the AUTOREG Procedure, 3rd Annual Conference of the Western Users of SAS Software for an excellent introduction to ARCH/GARCH methodology.

Following is the result of using the SAS PROC AUTOREG stepwise procedure to test for autocorrelation. The results of Godfrey's serial correlation test indicate that a high degree of autocorrelation exists in the MEI series.

Using Stepwise AUTOREG option

Dependent Variable = MEI

Ordinary Least Squares Estimates

SSE	521.9699	DFE	583
MSE	0.895317	Root MSE	0.946212
SBC	1606.21	AIC	1597.467
Reg Rsq	0.1403	Total Rsq	0.1403
Durbin-Watson	0.1499		

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	500.2182	0.0001
AR(+ 2)	500.7407	0.0001
AR(+ 3)	501.2850	0.0001
AR(+ 4)	503.7191	0.0001
AR(+ 5)	503.8035	0.0001
AR(+ 6)	504.1319	0.0001
AR(+ 7)	505.0320	0.0001
AR(+ 8)	505.5365	0.0001
AR(+ 9)	505.5640	0.0001
AR(+ 10)	505.6173	0.0001
AR(+ 11)	505.6445	0.0001
AR(+ 12)	505.8811	0.0001

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-0.314771	0.0558	-5.637	0.0001
DATE	1	0.000074231	7.611E-6	9.753	0.0001

Estimates of Autocorrelations

```
Lag  Covariance  Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0    0.892256     1.000000 |                    |********************|
  1    0.823859     0.923344 |                    |******************  |
  2    0.751164     0.841870 |                    |*****************   |
  3    0.674491     0.755939 |                    |***************     |
  4    0.583675     0.654156 |                    |*************       |
  5    0.497713     0.557814 |                    |***********         |
  6     0.41282     0.462669 |                    |*********           |
  7    0.322297     0.361216 |                    |*******             |
  8    0.232304     0.260356 |                    |*****               |
  9    0.149351     0.167386 |                    |***                 |
 10    0.072022     0.080719 |                    |**                  |
 11    0.007879     0.008831 |                    |                    |
 12     -0.0519    -0.058168 |                   *|                    |
 13    -0.10209    -0.114423 |                  **|                    |
```

Backward Elimination of Autoregressive Terms

Lag	Estimate	t-Ratio	Prob
2	0.002625	0.0453	0.9638
9	-0.007010	-0.1207	0.9040
13	0.006223	0.1498	0.8810
7	0.020127	0.3475	0.7283
5	-0.030821	-0.5376	0.5911
10	0.045763	0.9246	0.3555
11	-0.044914	-0.9396	0.3478
12	0.023731	1.0878	0.2771
6	-0.051436	-1.2485	0.2123
3	-0.079294	-1.6048	0.1091
4	0.044989	1.4769	0.1402

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.95374030	0.016723	-57.032
8	0.08415031	0.016723	5.032

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.9224
2	0.8401
3	0.7533
4	0.6627
5	0.5687
6	0.4717
7	0.3722
8	0.2709

Maximum Likelihood Estimates

SSE	71.53375	DFE	581
MSE	0.123122	Root MSE	0.350887
SBC	458.5308	AIC	441.0443
Reg Rsq	0.0174	Total Rsq	0.8822
Durbin-Watson	2.0130		

Godfrey's Serial Correlation Test

Alternative	LM	Prob>LM
AR(+ 1)	0.7008	0.4025
AR(+ 2)	1.3227	0.5162
AR(+ 3)	1.5983	0.6598
AR(+ 4)	3.0370	0.5517
AR(+ 5)	3.2720	0.6581
AR(+ 6)	3.2741	0.7737
AR(+ 7)	4.1700	0.7600
AR(+ 8)	4.5901	0.8004
AR(+ 9)	4.8731	0.8452
AR(+ 10)	7.3103	0.6959
AR(+ 11)	7.9150	0.7209
AR(+ 12)	8.4073	0.7525

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-0.297870	0.1537	-1.938	0.0531
DATE	1	0.000066916	0.000021	3.200	0.0014
A(1)	1	-0.959716	0.0166	-57.878	0.0001
A(8)	1	0.093760	0.0168	5.571	0.0001

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.9260
2	0.8451
3	0.7579
4	0.6650
5	0.5672
6	0.4651
7	0.3595
8	0.2513

Autoregressive parameters assumed given.

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-0.297870	0.1536	-1.939	0.0530
DATE	1	0.000066916	0.000021	3.204	0.0014

Testing for Heteroscedasticity

In order to test for heteroscedasticity with PROC AUTOREG, the ARCHTEST option is used in the model statement. The resulting small p-values for the Q statistics test and the Lagrange multiplier (LM) test indicate that a long memory period GARCH model may be the appropriate model to use.

Dependent Variable = MEI

Ordinary Least Squares Estimates

SSE	521.9699	DFE	583
MSE	0.895317	Root MSE	0.946212
SBC	1606.21	AIC	1597.467
Reg Rsq	0.1403	Total Rsq	0.1403
Durbin-Watson	0.1499	PROB<DW	0.0001

Q and LM Tests for ARCH Disturbances

Order	Q	Prob>Q	LM	Prob>LM
1	459.815	0.0001	457.999	0.0001
2	789.957	0.0001	461.099	0.0001
3	999.684	0.0001	463.998	0.0001
4	1126.34	0.0001	464.009	0.0001
5	1196.70	0.0001	464.077	0.0001
6	1233.88	0.0001	464.100	0.0001
7	1251.01	0.0001	464.185	0.0001
8	1255.90	0.0001	465.110	0.0001
9	1256.29	0.0001	465.118	0.0001
10	1256.31	0.0001	466.219	0.0001
11	1256.43	0.0001	466.541	0.0001
12	1256.46	0.0001	467.296	0.0001

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-0.314771	0.0558	-5.637	0.0001
DATE	1	0.000074231	7.611E-6	9.753	0.0001

Estimates of Autocorrelations

```
Lag  Covariance  Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  0    0.892256     1.000000 |                    |********************|
  1    0.823859     0.923344 |                    |******************  |
  2    0.751164     0.841870 |                    |*****************   |
  3    0.674491     0.755939 |                    |***************     |
  4    0.583675     0.654156 |                    |*************       |
  5    0.497713     0.557814 |                    |***********         |
  6     0.41282     0.462669 |                    |*********           |
  7    0.322297     0.361216 |                    |*******             |
  8    0.232304     0.260356 |                    |*****               |
```

Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.95374030	0.016723	-57.032
8	0.08415031	0.016723	5.032

Expected Autocorrelations

Lag	Autocorr
0	1.0000
1	0.9224
2	0.8401
3	0.7533
4	0.6627
5	0.5687
6	0.4717
7	0.3722
8	0.2709

Maximum Likelihood Estimates

SSE	71.53375	DFE	581
MSE	0.123122	Root MSE	0.350887
SBC	458.5308	AIC	441.0443
Reg Rsq	0.0174	Total Rsq	0.8822
Durbin-Watson	2.0130	PROB<DW	0.5464

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	-0.297870	0.1537	-1.938	0.0531
DATE	1	0.000066916	0.000021	3.200	0.0014
A(1)	1	-0.959716	0.0166	-57.878	0.0001
A(8)	1	0.093760	0.0168	5.571	0.0001

```
Expected Autocorrelations
Lag Autocorr
  0   1.0000
  1   0.9260
  2   0.8451
  3   0.7579
  4   0.6650
  5   0.5672
  6   0.4651
  7   0.3595
  8   0.2513
```

```
Autoregressive parameters assumed given.
Variable  DF      B Value  Std Error  t Ratio Approx Prob
Intercept  1    -0.297870     0.1536   -1.939     0.0530
DATE       1  0.000066916   0.000021    3.204     0.0014
```

Estimating GARCH(1,1)

```
Dependent Variable = MEI
Ordinary Least Squares Estimates
SSE           521.9699    DFE              583
MSE           0.895317    Root MSE    0.946212
SBC            1606.21    AIC         1597.467
Reg Rsq         0.1403    Total Rsq     0.1403
Durbin-Watson   0.1499
```

```
Variable  DF    B Value  Std Error  t Ratio Approx Prob
Intercept  1  -0.314771     0.0558   -5.637    0.0001
DATE       10.000074231    7.611E-6    9.753    0.0001
```

```
Estimates of Autocorrelations
Lag  Covariance  Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  0    0.892256    1.000000 |                    |********************|
  1    0.823859    0.923344 |                    |******************  |
  2    0.751164    0.841870 |                    |*****************   |
  3    0.674491    0.755939 |                    |***************     |
  4    0.583675    0.654156 |                    |*************       |
  5    0.497713    0.557814 |                    |***********         |
  6     0.41282    0.462669 |                    |*********           |
  7    0.322297    0.361216 |                    |*******             |
  8    0.232304    0.260356 |                    |*****               |
```

```
Estimates of the Autoregressive Parameters
  Lag  Coefficient    Std Error    t Ratio
    1  -0.95374030     0.016723    -57.032
    8   0.08415031     0.016723      5.032
```

GARCH Estimates

```
SSE             72.08725    OBS              585
MSE             0.123226    UVAR        0.127878
Log L           -129.529    Total Rsq     0.8813
SBC             303.6592    AIC          273.058
Normality Test  2057.552    Prob>Chi-Sq   0.0001
```

```
Variable  DF    B Value    Std Error   t Ratio  Approx Prob
Intercept  1  -0.258135       0.0940    -2.747       0.0060
DATE       10.000056306     0.000015     3.876       0.0001
A(1)       1  -0.979787       0.0170   -57.704       0.0001
A(8)       1   0.108132       0.0147     7.356       0.0001
ARCH0      1   0.019727      0.00585     3.370       0.0008
ARCH1      1   0.280158       0.0706     3.966       0.0001
GARCH1     1   0.565580       0.0993     5.693       0.0001
```

Conclusion

The focus of this poster was to illustrate how to detect and correct for the statistical problem of heteroscedasticity in the MEI time series. PROC AUTOREG was used to detect and estimate the type and order of the ARCH/GARCH process. As judged by the traditional summary statistics, it appears that a GARCH (1,1) procedure adequately models the MEI series.

Contact Information

Joseph Earley
Department of Economics
Loyola Marymount University
7900 Loyola Blvd.
Los Angeles, CA 90045-8410
Phone: (310) 338-4572
e-mail: jearley@lmumail.lmu.edu

References

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994), Time Series Analysis: Forecasting and Control, Third Edition, Englewood Cliffs, NJ: Prentice Hall.

Choi, ByouongSeon (1992), ARMA Model Identification, New York: Springer-Verlag. Engle, R.F. (1982) "Autoregressive Conditional Heteroscedasticity Estimates of the Variance of United Kingdom Inflation," Econometrica, 50, 987-1007.

Bollerslev, T. (1986) "Generalized Autoregressive Heteroskedasticity," Journal of Econometrics, 31, 307-327.

Pankratz, Alan (1983), Forecasting with Univariate Box-Jenkins Models: Concepts and Cases, New York, John Wiley & Sons.

SAS Institute Inc., SAS/ETS User's Guide, Version 6, Second Edition, Cary, NC: SAS Institute Inc., 1988.

SAS Institute Inc., SAS/ETS Software: Applications Guide 1, Version 6, First Edition, 1991.

SAS Institute Inc., SAS/ETS Software: Applications Guide 2, Version 6, First Edition, 1993.

SAS Institute Inc., SAS/STAT User's Guide, Vols I and II, 1990 Version 6, Fourth Edition. Cary, N.C.: SAS Institute Inc.

Wei, William (1990), Time Series Analysis: Univariate and Multivariate Methods, Redwood City, CA: Addison-Wesley Publishing Company Inc.

Woodward, Donna (1995) An Introduction to ARCH/GARCH Modeling Using the AUTOREG Procedure, 3rd Annual Conference of the Western Users of SAS Software.

SAS Code for ARCH/GARCH Model:

```
proc autoreg data = a;
   model mei = date /  nlag = (1  8)
     garch =  (q =1, p = 1)  maxit = 50;
     output  out = out  cev = vhat;
run;
```