Journal of the Nigerian Association of Mathematical Physics
Volume **11 (November 2007), 621 - 626**

Identification of the time series interrelationships with reference to dynamic regression models

N. P. Olewuezi
Department of Mathematics and Computer Science
Federal University of Technology, Owerri, Nigeria.
e-mail: ngolewe@yahoo.com. 08034933133

Abstract

In this study, the model of interest is that of a rational distributed lag function Y on X plus an independent Autoregressive Moving Average (ARMA) model. To investigate the model structure relating X and Y we considered the inverse cross correlation function for the observed and residual series in the presence of outliers. A two stage identification procedure is presented which involves fitting univariate time series model to each series and identifying a dynamic shock model relating the two univariate model series. The models so far obtained were combined to identify a dynamic regression model, which were fitted in the usual ways. From our findings, there was a reduction in the error variance of the final model with the outlier free stationary series which is an indication that the two-stage procedure is reliable and efficient.

***Keywords*:** dynamic shock model, inverse cross correlation, rational distributed lag function, outliers, dynamic regression model.

1.0 Introduction

In identifying regression models relating two time series, a methodology is introduced. If it is felt that *X* leads *Y*, then one may attempt to build a dynamic regression model relating the two series. By a dynamic regression model or distributed lag model we mean a regression of y_t, *y* at time *t*, on the present and past values *x*, $x_s(s \leq t)$. The model form of primary interest is that of *y* on *x* plus an independent noise term of the general mixed autoregressive moving average type. To identify the relationship between the two series *X* and *Y*, Haugh and Box (1977) [5] characterized each of their univariate models separately and the relationship between the two univariate residual series driving each time series. At the first stage, an autoregressive integrated moving average processes are fitted to each of *X* and *Y* series. The residual series U_x and U_y from these fits are then inverse cross correlated $r_{iu_x u_y}$ thereby identifying a tentative dynamic shock model which relates U_x to U_y. By recombining the two univariate models for *X* and *Y* with the identified model connecting U_x and U_y a distributed lag model relating X to Y may be identified, fitted and checked using the methodology by Box and Jenkins (1970) [1]. This method can be used for both the outlier free (OF) and outlier contaminated (OC) series and any outlying observations found are removed accordingly.

2.0 The inverse cross-correlation estimator

The cross-correlation function $\rho_{xy}(\cdot)$ at lag k is defined as $\rho_{xy}(k) = \frac{E[(x_t - \mu_x)(y_{t+k} - \mu_y)]}{\sigma_x \sigma_y}$ where μ and σ denote the mean and standard deviation. Let y_{1t} and y_{2t} be stationary series such that $E(y_{jt}) = \mu_j$, $(j = 1,2)$ and $Var(y_{jt}) = \gamma_j = \sigma_j^2 \, (j = 1,2)$. Then the cross covariance between y_{1t} and $y_{2,t-k}$ is defined as $Cov(y_{1t}, y_{2,t-k}) = E[y_{1t}, y_{2,t-k}] - \mu_1 \mu_2 = \gamma_{12}(k)$. Oluwuezi and Shangodoyin (2005) [6] defined the inverse auto-covariance function at lag k as

$$\gamma_i(k) = 2\pi \sum_{k=-\infty}^{\infty} \left(\frac{1}{k}\right)\left(\frac{1}{\gamma_j}\right) \sin 2\pi k \qquad (2.1)$$

which was expressed as

$$\gamma_i(k) = \begin{cases} \frac{2\pi^2}{\gamma_0}, & k = 0 \\ 4\pi \sum_{j=1}^{k} \left(\frac{1}{j}\right)\left(\frac{1}{\gamma_k}\right) \sin 2\pi j, & k \neq 0 \end{cases}$$

where γ_k is the auto-covariance function at lag k and γ_0 is the variance of the process. The inverse cross correlation function is given by $\rho_{i,12}(k) = \frac{\gamma_{i12}(k)}{\sigma_1 \sigma_2} = \frac{2\pi^2}{\gamma_0} + 4\pi \sum \tau_k \sin 2\pi k$, where $\tau_k = \frac{1}{k\gamma_{12}(k)\sigma_1\sigma_2}$.

3.0 The inverse co-variance structure

Let X and Y be jointly covariance stationary time series. For any additive outlier series we have

$$\left.\begin{aligned} X_t &= V(B)U_{x_t} + \alpha_T \xi_t^{(T)} \\ Y_t W(B)U_{y_t} &+ \beta_T \xi_t^{(T)} \end{aligned}\right\} \qquad (3.1)$$

where α and β are the estimates of the outliers at time $T = t$, for the two series respectively. From equation (3.1), the inverse covariance structure of X and Y defined since the innovations of the specification (X_t, Y_t) are independent may be written as

$$\left.\begin{aligned} \gamma_{i_x}(k) &= V(B)^2 \gamma_{iu}(k)_x + \gamma_{i\alpha}^*(k) \\ \gamma_{iy}(k) &= W(B)^2 U \gamma_{iu_y}(k) + \gamma_{i\beta}^*(k) \\ \gamma_{ixy} &= V(B)W(B)\gamma_{iu_x u_y} + \gamma_{i\alpha\beta}^*(k) \end{aligned}\right\} \qquad (3.2)$$

where

$$\gamma_{iu_j}(k) = 0 \text{ for all } k \neq 0; \; j = x, y$$

$$\gamma_{iu_j}(0) = \sigma_{uj}^2 \text{ for all } k = 0$$

$$\gamma_i^*(.) \text{ exists for all } t = T \text{ and zero}$$

Otherwise, using equation (2.1), the inverse covariance structures become

$$\gamma_{ix}(k) = V(B)^2\left[\frac{2\pi^2}{\gamma_0} + 4\pi\sum_{k=1}^{\infty}\frac{\sin 2\pi k}{k\gamma_{ux}(k)} + \gamma_{i\alpha}^{*}(k)\right]$$

$$\gamma_{iy}(k) = W(B)^2\left[\frac{2\pi^2}{\gamma_0} + 4\pi\sum_{k=1}^{\infty}\frac{\sin 2\pi k}{k\gamma_{uy}(k)} + \gamma_{i\beta}(k)\right]$$

and
$$\gamma_{ixy}(k) = V(B)W(B)\left[\frac{2\pi^2}{\gamma_0} + 4\pi\sum_{k=1}^{\infty}\frac{\sin 2\pi k}{k\gamma_{u_x u_v}(k)} + \gamma_{i\alpha\beta}^{*}(k)\right]$$

Hence the inverse cross correlation estimator is derived as

$$\rho_{i12}(k) = \frac{2\pi^2}{\gamma_0} + 4\pi\sum_{k=1}^{\infty}\tau_k \sin 2\pi k$$

where
$$\tau_k = \frac{1}{k\gamma_{12}\sigma_1\sigma_2}$$

4.0 Identification of the joint univariate model

Using Haugh L.D. and Box G.E.P. (1977) approach we define for the OF white noise process

$$U'_t = \begin{bmatrix} U'_{xt} \\ U'_{yt} \end{bmatrix} = U(B)a'_t$$

$$U(B) = \begin{bmatrix} U_X(B) & U_{xy}(B) \\ U_{yx}(B) & U_y(B) \end{bmatrix}$$

$$a'_t = \begin{bmatrix} a'_{x_t} \\ a'_{y_t} \end{bmatrix}$$

We define
$$U_{jj}(B) = \theta_{jj}(B)\,\phi_{jj}(B)^{-1}$$

and
$$U_{j1}(B) = \theta_{j1}(B)\,\phi_{j1}(B)^{-1}$$

for
$$j1 = x, y \ and \ \theta_{xy}(0) = 0$$

At negative lags when no significant inverse cross correlation occurs between U_x and U_{y_1} we now form a complete OF model for *X* and *Y*. If we assume that there are no feedback effect in that no significant inverse cross correlation occurs at negative lags and that $U_{xy}(B) = 0, U_x(B) = 1$ and $a'_{y_t} = a'_t$ then the full model relating X and Y could be written out for additive and innovative outlier models as

$$Y_t = V_Y(B)\,U_{YX}(B)\,V_X(B)^{-1}X_t - V_X(B)^{-1}D_{XA,T} + D_{YA,T} + V_Y(B)\,U_Y(B)a'_t$$

and

$$Y_t = V_Y(B)\,U_{YX}(B)\,V_X(B)^{-1}X_t - V_X(B)^{-1}V_{Xt-T}DX_{1,T} + V_{Yt-T}D_{Y1,T} + V_Y(B)\,U_Y(B)a'_t$$

Let us now illustrate the relative merit of identifying time series interrelationships using inverse cross correlation function for the observed and residual series in the presence of outliers. The comparison is performed with respect to the model residual variance and the diagnosing test statistic.

5.0 Illustration

To illustrate the feasibility of the method, three series are used. Series A is the stack-loss data [Brownlee (1965)], Series B is the Gas Furnace data [Box and Jenkins (1976)] Series C is the simulated data of size 100.

A two-stage identification procedure is presented which involves fitting univariate time series models to each series and identifying a dynamic shock model relating the two univariate models which are combined to identify a dynamic regression model. We assumed an additive outlier models and used Tsay (1986) testing criteria for outlier detection.

5.1. Series A

The autocorrelation function (ACF) and its inverse (IACF) suggested an AR(1) for both the OC and OF series. We fitted these models.

5.2 OC Series

$$(1-0.9854B)\ X_t = U_{x_t} \left(\sigma^2_{u_{xt}} = 77.01\right)$$

and

$$(1-0.92398B)\ Y_t = U_{y_t} \left(\sigma^2_{u_{yt}} = 34.89\right)$$

5.3 OF Series

$$(1-0.9898B)\ X_t = U_{x_t} \left(\sigma^2_{u_{xt}} = 24.79\right)$$

and

$$(1-0.9376B)\ Y_t = U_{y_t} \left(\sigma^2_{u_{yt}} = 11.93\right)$$

A reduction of about 67.81% and 65.81% are achieved for both X and Y series respectively with the OF series. At the second stage, with $\rho_{iu_xu_y} \neq 0$ for k = 0 and 1 leads to the dynamic shock models given as

5.4 OC Series

$$U_{y_t} = (1-0.8742B)\ U_{x_t}$$

With no outlier found in either U_{x_t} or U_{y_t} series

5.5 OF Series

$$U_{y_t} = (1-0.6528B)\ U_{x_t}$$

5.6 Completing identification

A substitution of the identified univariate models for *X* and *Y* into the preceding relationship leads to the dynamic regression models of the form

5.7 OC Series

$$Y = (1-1.8596B)(1-0.923B)^{-1} X_t + (1-0.4938B)(1-0.923B)^{-1} a_t$$

with $\sigma^2_{a_t} = 12.39$

5.8 OF Series

$$Y_t = (1-1.6428B)(1-0.9376B)^{-1} X_t + (1-0.7010B)(1-0.9376B)^{-1} a_t$$

The error variance of the final model is reduced by 55.77% with the OF stationary series.

5.91 Series B

The ACF and IACF suggested an AR(1) for the X series and for the first difference to the Y series. We fitted these models.

5.10 OC Series

$$(1-0.0612B)X_t = U_{x_t};\ (\sigma^2_{u_{xt}} = 41.58)$$

and

$$(1-0.09867B)Y_t = U_{x_t}\ \ (\sigma^2_{u_{yt}} = 78.48)$$

5.11 OF Series

$$(1-0.9490B)X_t = U_{x_t} \;\; (\sigma^2_{u_{xt}} = 0.114)$$

and $$(1-0.8206B)Y_t = U_{y_t} \;\; (\sigma^2_{u_{yt}} = 0.1839)$$

A reduction of about 99.72% for the X series and about 99.76% for the Y series were recorded in the residual variance of the OF models. At the second stage with $\rho_{iu_x u_y} \neq 0$ for k = 0 and 1 leads to the dynamic shock models $U_{y_t} = (1+0.866B)\, U_{x_t}$

With no outlier found in either U_{x_t} or U_{y_t} series

T = 28, 33, 45, 62, 190, 203, 464.

and T = 21, 60, 119, 226, 240

For the OF series we have $U_{y_t} = (1+0.7915B)\, U_{x_t}$

5.12 Completing identification

The model

$$Y = (1+0.804\,B)(1-0.9867\,B)^{-1} X_t + (1+0.756\,B)(1-0.9867\,B)^{-1} a_t$$

with $$\sigma^2_{a_t} = 12.31$$

was obtained for the OC stationary series and the model for the OF stationary series give us

$$Y_t = (1+0.1575B)(1-0.8206B)^{-1} X_t + (1-0.6530B)(1-0.8206B)^{-1} a_t$$

with $$\sigma^2_{a_t} = 0.10$$

The error variance of the final model is reduced by 99.19% with the OF series.

5.13 Series C

We fitted an ARMA (1,1) for the outlier stationary series X and Y. We then fitted the following models for both the OC and OF models.

5.14 OC Series

$$(1+0.9898B)X_t = (1+0.9724B)U_{x_t} (\sigma^2_{u_{xt}} = 106.713)$$

and $$(1+0.0561B)Y_t = (1-0.9692B)U_{y_t} (\sigma^2_{u_{yt}} = 3.5097)$$

5.15 OF Series

$$(1+0.0085B)X_t = (1-0.9824B)\, U_{x_t}; (\sigma^2_{u_{xt}} = 0.9667)$$

and $$(1+0.8345B)Y_t = (1-0.980B)U_{y_t} (\sigma^2_{u_{yt}} = 0.9484)$$

A reduction of about 99.09% for the X series and about 72.98% for the Y series were recorded in the residual variance of the OF models. At the second stage with $\rho_{iu_x u_y} \neq 0$ for k = 0 and 1 leads to the dynamic shock models

5.16 OC Series

$$(1+0.0617B)\, U_{y_t}; = (1-0.7593B)\, U_{x_t}$$

With no outlier found in U_{x_t} or U_{y_t} series

5.17 OF Series

$$(1+0.9215B) = U_{y_t} \;\; (1-0.6320B)\, U_{x_t}$$

5.18 Completing identification

The model

$$Y_t = (1-1.7728B)(1+0.0561B)^{-1} + (1-0.1914B)(1+0.0561B)^{-1} a_t$$

with $\sigma^2_{a_t} = 8.936$ is obtained for the OC stationary series and for the OF we have

$$Y_t = (1 - 1.5426B)(1 - 0.8345B)^{-1} X_t + (1 - 0.329B)(1 - 0.8345B)^{-1} a_t$$

with $\sigma^2_{a_t} = 0.831$

The error variance of the final model is reduced by 90.70% with the OF stationary series.

6.0 Conclusion

This study revealed that once U_x and U_y have been obtained at the first stage of identification, the bivariate dynamic shock identification which follows depends in no way on the univariate models employed. There is a reduction in the error variance of the final with the OF stationary series which is an indication that the two stage procedure is reliable and efficient. A problem deserving future investigation is the distribution of the residual inverse cross correlation function when the series are not independent.

Acknowledgement

I humbly acknowledge the assistance of Dr. D.K. Shangodoyin for his immense contribution towards my academic attainment.

References

[1] G.E.P Box and G.M. Jenkins, Time Series Analysis, Forecasting and Control, Holden Day, San Francisco, 1976.
[2] C. Chatfield, Inverse Autocorrelations, J. Roy Stat. Ser. A. 142 (1979) 363-377.
[3] N.R. Draper and H. Smith, Applied Regression Analysis, Wiley, New York, 1966.
[4] A. J. Fox, Outliers in Time Series, J. Roy Stat. Soc. Ser. B. 34 (1972), 350-363.
[5] L.D. Haugh and G.E.P. Box, Identification of Dynamic Regression (Distributed Lag) Models Connecting [6] Two Time Series, J. Amer. Stat Asso. (March 1977) vo. No. 357, 121-131.
[7] N. P. Olewuezi and D.K. Shangodoyin, Contributions of Inverse Autocorrelation Function in Transfer Function Modeling, Far East J. Theo Stat. 17 (2) (2005) 227-238.
[8] D.K. Shangodoyin, On the Specification of Time Series models in the presence of aberrant Observations, Ph.D. Thesis, Dept of Statistics, Univ. of Ibadan, Nigeria (1994).
[9] R. S. Tsay, Time Series model specification in the presence of outliers, J. AMer. Stat. Asso. 81 (1986), 132-141.