University of Wollongong
Research Online

Applied Statistics Education and Research Collaboration (ASEARC) - Conference Papers

Faculty of Engineering and Information Sciences

2011

Threshold Autoregressive Models in Finance: A Comparative Approach

David Gibson
University of Newcastle

Darfiana Nur
University of Newcastle

Publication Details

Gibson, David; and Nur, Darfiana, Threshold Autoregressive Models in Finance: A Comparative Approach, Proceedings of the Fourth Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Threshold Autoregressive Models in Finance: A Comparative Approach

Abstract
Financial instruments are known to exhibit abrupt and dramatic changes in behaviour. This paper investigates the relative efficacy of two-regime threshold autoregressive (TAR) models and smooth threshold autoregressive (STAR) models, applied successfully to econometric dynamics, in the finance domain. The nature of this class of models is explored in relation to the conventional linear modeling approach, with reference to simulated data and real stock return indices.

Keywords
Threshold, Nonlinear, Autoregressive, STAR

Publication Details
Gibson, David; and Nur, Darfiana, Threshold Autoregressive Models in Finance: A Comparative Approach, Proceedings of the Fourth Annual ASEARC Conference, 17-18 February 2011, University of Western Sydney, Paramatta, Australia.

This conference paper is available at Research Online: http://ro.uow.edu.au/asearc/26

Threshold Autoregressive Models in Finance: A Comparative Approach

David Gibson and Dr Darfiana Nur

The University of Newcastle, Callaghan, NSW, 2308, AUSTRALIA

Abstract

Financial instruments are known to exhibit abrupt and dramatic changes in behaviour. This paper investigates the relative efficacy of two-regime threshold autoregressive (TAR) models and smooth threshold autoregressive (STAR) models , applied successfully to econometric dynamics, in the finance domain. The nature of this class of models is explored in relation to the conventional linear modeling approach, with reference to simulated data and real stock return indices.

Key words: Threshold, Nonlinear, Autoregressive, STAR

1. Introduction

Autoregressive models have been applied across diverse fields of endeavour. Yule (1927) applied the first autoregressive model to the understanding of Wolfer's sunspot numbers over time, but authors such as Pesaran and Timmerman (1995) have extended the autoregressive model into the financial domain.

Practitioners in many fields are increasingly faced with real data possessing nonlinear attributes. It is known that stationary Gaussian autoregressive models are structurally determined by their first two moments. Consequently, linear autoregressive models must be time reversible. Many real datasets are time irreversible, suggesting that the underlying process is nonlinear. Indeed, in Tong's seminal paper on threshold models, he would argue that no linear Gaussian model could explain the cyclical dynamics observed in sections of the lynx data (Tong and Lim, 1980). Furthermore, he argued that characteristics of nonlinear models, such as time irreversibility and limit cycles, mandated the development of practical nonlinear models to help resolve ongoing difficulties in real data. Tong's explanation and application of locally linear threshold models introduced striking opportunities for model building strategies.

2. Threshold Autoregressive (TAR) Models

The *Threshold Autoregressive* (TAR) family proposed and explained by Tong (1983) are contained within the state-dependent (regime-switching) model family, along with the bilinear and exponential autoregressive (EAR) models.

The simplest class of TAR models is the *Self Exciting Threshold Autoregressive* (SETAR) models of order p introduced by Tong (1983) and specified by the following equation:

$$Y_t = \begin{cases} a_0 + \sum_{j=1}^{p} a_j Y_{t-j} + \epsilon_t & \text{if } Y_{t-d} \leq r \\ (a_0 + b_0) + \sum_{j=1}^{p} (a_j + b_j) Y_{t-j} + \epsilon_t. & \text{if } Y_{t-d} > r \end{cases} \tag{1}$$

TAR models are piecewise linear. The threshold process divides one dimensional Euclidean space into k regimes, with a linear autoregressive model in each regime. Such a process makes the model nonlinear for at least two regimes, but remains locally linear (Tsay, 1989). One of the simplest of TAR models equates the state determining variable with the lagged response, producing what is known as a *Self-Exciting Threshold Autoregressive* (SETAR) model.

A comparatively recent development is the *Smooth Transition Autoregressive* (STAR) model, developed by Terasvirta and Anderson (1992). The STAR model of order p model is defined by

$$\begin{aligned} Y_t &= a_0 + a_1 Y_{t-1} + \ldots + a_p Y_{t-p} + \\ &(b_0 + b_1 Y_{t-1} + \ldots b_p Y_{t-p}) G\left(\frac{Y_{t-d} - r}{z}\right) + \varepsilon_t \end{aligned} \tag{2}$$

where d, p, r, $\{\varepsilon_t\}$ are as defined above, z is a smoothing parameter $z \in \Re^+$ and G is a known distribution function which is assumed to be continuous. Transitions are now possible along a continuous scale, making the regime-switching process 'smooth'. This helps overcome the abrupt switch in parameter values characteristic of simpler TAR models.

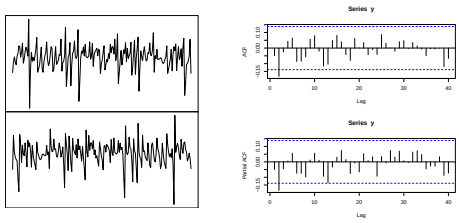

Figure 1: Original data and reversed data (left) and sample autocorrelation (ACF) and partial autocorrelation (PACF) (right)

3. Estimation via Conditional Least Squares (CLS)

The most popular method for estimation is conditional least squares (CLS). In this approach, the predictive sum of squared errors are minimised to obtain parameter estimates. Firstly, let $E|Y_t| < \infty, t = 1, 2, ...$ and $\Theta = (a_0, a_1, ..., a_p, b_0, b_1, ..., b_p, \sigma^2)$. This gives the necessary information for calculating the conditional least squares, and producing an estimate for Θ can be accomplished by minimising the residual sum of squares such that:

$$Q_N(\Theta) = \sum_{j=1}^{N} [Y_j - E_\theta(Y_j|B_{j-1})]^2 \tag{3}$$

with respect to Θ.

4. Simulation Study

This simulation is the realisation of a simple two-regime SETAR model produced to allow for the identification of nonlinear phenomena and a sample analytical process. Coefficient vectors in the lower and upper regime have been set at $[0, 0.5]$ and $[0, -1.8]$ respectively, with $n = 200$, threshold parameter $= -1$, delay = 1, and the noise standard deviations in the lower and upper regime at 1 and 2. The specific two-regime model form in this case is:

$$Y_t = \begin{cases} 0.5Y_{t-1} + \epsilon_t & \text{if } Y_{t-1} \leq -1 \\ \\ -1.8Y_{t-1} + 2\epsilon_t & \text{if } Y_{t-1} > -1 \end{cases}$$

The characteristics of nonlinearity are present here. Asymmetry is evident in the rise and fall of the series and the movement of the series suggest limit cycle behaviour. The series is also time irreversible, as demonstrated in the reversed plot. The sample autocorrelation function (Figure 1) reveals a significant second lag, with all other lags within the confidence interval. In the PACF series, it is significant at the second lag and no other. The series satisfies stationarity conditions.

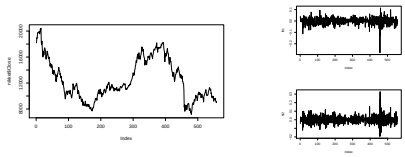

Figure 2: Sequence plot for the Nikkei-225 Index and First difference of the logged Nikkei-225 data

4.1. Identification and Estimation

An autoregressive moving-average (ARMA) with lags 1 and 1 respectively (ARMA(1,1)) was fitted.

```
Call:
arima(x = y, order = c(1, 0, 1))

Coefficients:
              ar1           ma1     intercept
coef    0.303654809 -6.058011e-01 -6.733345e-01
p-value 0.003634426  1.908790e-12  6.567002e-45

sigma2 estimated as 3.633:  log likelihood = -1052.67,  aic = 2111.35
```

Statistically significant autoregressive, moving-average and intercept coefficients indicate a general improvement in fit. The estimation of a SETAR (2,1,1) model produces the following output:

```
SETAR(2, 1 , 1 ) model delay = 1

            Estimate Std.Err t-value Pr(>|t|)
intercept-y  -0.0808  0.1825 -0.4427    0.659
lag1-y        0.4755  0.0535  8.8955    0.000

            Estimate Std.Err  t-value Pr(>|t|)
intercept-y   0.0905  0.2063   0.4385    0.662
lag1-y       -1.7153  0.1392 -12.3180    0.000
```

This is an edifying model estimation result, with the estimated and statistically significant values of the threshold at -1, the lower regime coefficient at 0.4755 and the upper coefficient at -1.7153. These point estimates are very close to the true values of 0.5 and -1.8 respectively. The fitted model may therefore be written as:

$$Y_t = \begin{cases} 0.4755Y_{t-1} + 1.0202\epsilon_t & Y_t \leq -1 \\ -1.7153Y_{t-1} + 1.9191\epsilon_t & Y_t > -1 \end{cases} \tag{4}$$

5. Diagnostic and Forecasting

Diagnostic procedures for the simulated data satisfied model assumptions. Both in-sample and out-of-sample forecasts revealed an improved fit for the SETAR model. Readers are referred to Gibson (2010) for a detailed exploration of these modeling aspects.

6. NIKKEI-225 Index Case Study

Weekly closing value data was obtained from the NIKKEI 225 Index, an average of 225 stocks traded on the Tokyo Stock Exchange, from January 2000 to September 2010. Characteristics of a nonlinear process seem to be present (as shown in Figure 2). Irregular

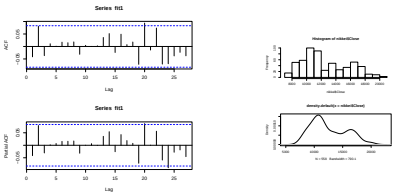

Figure 3: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the logged and differenced series and Histogram and density plot for the Nikkei-225 data

amplitude in the peaks and troughs suggest time irreversibility and asymmetry. Peaks rise and fall dramatically, and there appears to be a very slight downward trend. The first difference of the logged data (Figure 2) revealed an irregular spike around the 450th data point.

6.0.1. *Sample Correlation*

The logged and differenced series (Figure 3), both the ACF and PACF reveal no significant values until the 20th lag. Non-significant lags can be evidence towards nonlinearity. A histogram and density plot of the data suggest a bimodal distribution, a characteristic common to nonlinear processes.

6.0.2. *Linearity Testing*

Test	Tar-F Test	LR Test	Keenan Test
Statistic	1.472	35.913	10.398
p	< 0.01	0.143	< 0.01

Disagreement is noted between the tests over the nature of the process. Tsay's TAR-F test successfully rejects the null hypothesis of a linear process, while Chan's Likelihood Ratio test fails to do so. A possible reason for this is that Chan's test retains the greatest power when the alternative is the "true" model under consideration.

6.0.3. *Identification and Estimation*

Iterative linear model building strategies with varying autoregressive and moving-average parameters met with mixed results. After many attempts, an appropriate linear model was specified for baseline comparison.

```
Call:\\
arima(x = fit2, order = c(4, 0, 0))

Coefficients:
             ar1           ar2           ar3           ar4     intercept
coef    -8.394919e-01 -5.584072e-01 -4.074595e-01 -2.018603e-01 -4.997183e-05
p-value  1.053899e-90  5.544701e-27  3.859383e-15  1.148524e-06  9.203625e-01
```

All coefficients are strongly significant here in the AR(4) base model. The final chosen SETAR model had a threshold delay d of 0, and autoregressive order 2 in the lower and upper regimes. This gives:

```
SETAR(2, 2 , 2 ) model delay = 0

               Estimate Std.Err  t-value Pr(>|t|)
intercept-fit2  -0.0265  0.0016 -16.2387   0.0000
lag1-fit2       -0.3437  0.0412  -8.3466   0.0000
lag2-fit2       -0.1290  0.0391  -3.2986   0.0011

               Estimate Std.Err t-value Pr(>|t|)
intercept-fit2   0.0258  0.0022 11.9718    0e+00
lag1-fit2       -0.5495  0.0568 -9.6697    0e+00
lag2-fit2       -0.1768  0.0513 -3.4490    7e-04
```

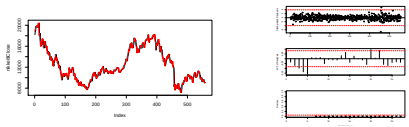

Figure 4: Fitted values for the AR(4) model and Residuals for the AR(4) model

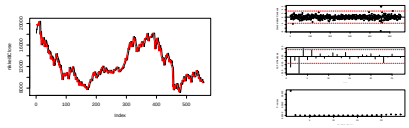

Figure 5: Fitted values for the SETAR model and Diagnostics for the SETAR model

All parameter estimates are significant in this model, with a considerable decrease in AIC. The SETAR (2,2,2) model obtained in the output above is:

$$Y_t = \begin{cases} -0.0265 - 0.3437Y_{t-1} - 0.1290Y_{t-2} + 0.0257\epsilon_t & Y_{t-1} \leq 0.00087 \\ \\ 0.0258 - 0.5495Y_{t-1} - 0.1768Y_{t-2} + 0.0309\epsilon_t & Y_t > -0.00087 \end{cases} \tag{5}$$

An additional procedure for automatic STAR model estimation was employed. This produced a satisfactory model, but integrated tests for regime addition beyond the first threshold were rejected.

6.0.4. *Diagnostics*

Residual plots of the linear model (Figure 4) reveal non-random residual scatter with distinct point compression near the centre, and autocorrelation from the 3rd to the 5th lag. Introducing the threshold value has improved model fit statistics (Figure 5), but there is little appreciable improvement in the fitted values relative to the original process. Standardised residuals demonstrate point dispersal at the two extremes. The sample ACF reveals significant lags 1, 3 and 4. The failure to reject the null hypothesis in a test for the inclusion of additional regimes has reverted the STAR model in this case to the simpler SETAR form.

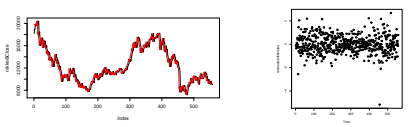

Figure 6: Fitted values for the STAR model and Residuals for the STAR model

Figure 7: Restricted in-sample forecast values for the AR(4) model and Restricted in-sample forecast values for the SETAR model

Figure 8: Restricted in-sample forecast values for the STAR model and Restricted out-of-sample forecast values for the AR(4) model

6.0.5. Forecasting

The fitted values for the AR(4) model fall centrally within the peak and trough of the original data, and is indicative of a reasonable model fit (see Figure 7). The SETAR model appears to be having difficulty in accurately predicting the final few values. A similar outcome is noted for the STAR model (Figure 8), with the predictions unable to suitably account for the drop in returns in the final few values. The out-of-sample predictions from the linear model reveal a slow downward trend as the series progresses. The SETAR model forecasts could be interpreted as an improvement on the ARIMA model, as shown in Figure 9. Rising values might indicate an attempt by the model to more effectively capture the volatility in the series, and reflect overall movements in the process. STAR model prediction values resemble strongly those seen in the SETAR model.

7. Conclusion

The extension of the autoregressive model to the regime-switching class is a natural progression, but the characteristics of nonlinearity are not always immediately detectable in visual plots such as the sequence chart. Within the two-regime Self-Exciting Threshold Autoregressive (SETAR) model simulation series these attributes are brought to the forefront and allow for the relative strengths of this class of models to be quantified. Exploratory analyses reveal with great speed the violation of linear modeling assumptions and the difficulties inherent to fitting this type of model. Data asymmetry and time irreversibility are traditional indicators of nonlinearity, while the rejection of normality in the data makes applying ARIMA models hazardous.

SETAR model building strategies are, conversely, ideal for this type of process. Model fitting summaries and diagnostic procedures reveal clear preferences, while both in-sample and out-of-sample forecasts are improved in the threshold model case. An unsurprising result, it suggests that real data sets demonstrating similar behaviour may benefit from the application of this model form. Empirical results from the application of the nonlinear models highlight the improvements in out-of-sample forecasting. Similar performance was also noted between the SETAR and STAR models in applying process dynamics to future data points. This outcome can be interpreted, however, as the result of smooth models with finite thresholds. In-sample forecast remains problematic, as in several cases the models were unable to properly replicate the observed model behaviour.

Figure 9: Restricted out-of-sample forecast values for the SETAR model and Restricted out-of-sample forecast values for the STAR model

References

[1] Gibson, David., (2010) "Threshold Autoregressive Models in Finance: A Comparative Approach", Master's Thesis, The University of Newcastle

[2] Pesaran, M Hashem, Timmerman, Allan., (1995) "Predictability of stock returns: Robustness and economic significance", The Journal of Finance, Vol. 50, pp 1201 - 1228

[3] Terasvirta, T, Anderson, H M., (1992) "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models", John Wiley & Sons, Vol. 7

[4] Tong, Howell., (1983) "Threshold Models in Non-linear Time Series Analysis", Springer-Verlag New York Inc.

[5] Tong, H., Lim, K.S., (1980) "Threshold autoregression, limit cycles, and cyclical data", Journal of the Royal Statistical Society, Vol. 42, pp 245 - 292

[6] Tsay, Ruey S., (1989) "Testing and modeling threshold autoregressive processes", Journal of the American Statistical Association, Vol. 84, pp 231 - 240

[7] Yule, Udny., (1927) "On a method for investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers", Philosophical Transactions of the Royal Society of London, Vol. 226, pp 267 - 298