Value-at-Risk Estimates from a SETAR Model

Joselito C. Magadia
School of Statistics
University of the Philippines Diliman

A self-exciting threshold autoregressive (SETAR) model will be fitted to PSEi and value-at-risk estimates would be computed. Backtesting procedures would be employed to assess the accuracy of the estimates and compared with estimates derived from two other approaches to VaR estimation.

Keywords: threshold models, backtesting, APARCH

I. Introduction

Given a sequence x_1, x_2,…, x_n observed at discrete time periods, modern approaches to the time series analyses of such data propose stochastic models of the form

$$X_t = f(X_{t-1}, X_{t-2}, ..., \varepsilon_t, \varepsilon_{t-1},…), \qquad (1)$$

wherein an observation x_t is represented by a random variable X_t and f is a measurable function of past representations of X and an independent innovation process.

By far, the most popular of such approaches are the ARIMA models of Box and Jenkins. Under conditions of stationarity and invertibility, an ARMA (p, q) model given by

$$X_t = \mu + \phi_1 X_{t-1} +…+\phi_p X_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + …+ \theta_q\varepsilon_{t-q}$$

can be represented in (possibly infinite order) autoregressive (AR) form as

$$X_t = \mu + \phi_1 X_{t-1}+\phi_2 X_{t-2} + … + \varepsilon_t , \varepsilon_t \sim iid(0,\sigma^2). \qquad (2)$$

A number of reasons could possibly account as to why researchers prefer to use this model. First, and foremost, is that the mathematical and statistical foundations for this model has almost been completely worked out, most specially the case wherein the error terms ε_t are assumed to be have a Gaussian distribution. In essence, this gives researchers the confidence that they are dealing with what seems to be a logically sound and valid model. Second, the stages in the model-building procedure, composed of the identification-estimation-diagnostics cycle proposed by its proponents can be routinely implemented. Working through just a score of textbook examples would be enough to acquaint an avid novice with the nuances that one could encounter with real-life data. The availability of statistical software that implements these procedures contributes immensely to reduce the computational workload required for most analyses. The estimation stage alone gives a number of methods to choose from. Graphs and plots to aid in the identification, post-hoc analysis and forecast performance are just a few mouse clicks (or command lines) away.

Perhaps another reason for choosing model (2) is due to the ease of interpretation that it offers. It simply suggests that any observation can be explained in terms of a linear combination of past observations plus an error term with a given variance. By way of extension, forecasts (and their margins of error) are just as easily estimated.

Linear models, however, do have limitations. Practitioners dealing with time series data often encounter phenomena which cannot be exhibited by any realization of these linear models. To name a few, Tong (1990) mentioned asymmetry, time-irreversibility, sudden bursts of very large amplitude at irregular time epoch, etc. Brock and Potter (1993) give a review of the presence of nonlinear behavior in macroeconomic and financial time series.

Any model, then, that does not conform to the linear form is a possible candidate for a nonlinear model. The reluctance to adopt just any form of nonlinearity is quite understandable since our experience in linear modeling imposes conditions on the structure and form of the model before it could be subjected to computer analytic methods.

Alternative models for consideration should not just be some arbitrary nonlinear combination of past values. A related field of study offers such a choice. Dynamical Systems Theory is a branch of mathematics that studies the behavior of a system's trajectory through time. Usually, a system's state at time (t + s) is given by a deterministic rule

$$X_{t+s} = f^k(X_t\,, X_{t-d}\,, \ldots, X_{t-(m-1)d}\,; \theta), \tag{3}$$

where

f^k is the k^{th} iterate of some generic function $f : \mathcal{R}^m \rightarrow \mathcal{R}$,

m is called the *embedding dimension*, and
d is called the *delay*.

Modifying (3) to account for randomness, one obtains (taking d =1),

$$X_t = f\,(X_{t-1}\,, X_{t-2}\,,\ldots, X_{t-m}, \varepsilon_t\,,\ldots, \varepsilon_{t-q}\,; \theta). \tag{4}$$

Tong (1990) provides an comprehensive discussion of a number of nonlinear time series models. Among them is a class of models collectively known as *threshold models*. This class includes regime-switching models as special cases. Among these *regime-switching models* is a subclass known as *self-exciting threshold AR (SETAR)* models. A SETAR (s, d; m_1, $m_2\ldots$, m_s) may be written as

$$X_t = \begin{cases} \phi_1^{(1)} X_{t-1} + \ldots + \phi_{m_1}^{(1)} X_{t-m_1} + \varepsilon_t^{(1)}, & r_0 < X_{t-d} \le r_1 \\ \phi_1^{(2)} X_{t-1} + \ldots + \phi_{m_2}^{(2)} X_{t-m_2} + \varepsilon_t^{(2)}, & r_1 < X_{t-d} \le r_2 \\ \vdots & \\ \phi_1^{(s)} X_{t-1} + \ldots + \phi_{m_s}^{(s)} X_{t-m_s} + \varepsilon_t^{(s)}, & r_{s-1} < X_{t-d} \le r_s \ , \end{cases} \tag{5}$$

with $\varepsilon_t^{(i)} \sim iid\left(0, \sigma_{(i)}^2\right), \sigma_{(i)}^2 < \infty$, r_0 = -∝ and r_s = ∝. In contrast to the more popular Markov switching models, wherein the regimes are determined by some discrete Markov state variable, the regimes in (5) are delineated by some lagged value of the time series itself. The primary purpose in suggesting this model is the ease of interpretation that it offers. It is a natural extension of the AR model and the values of the cut-off points could offer additional insight into the nature of time series. Another property of this class of models is that the trajectories of these models can exhibit behavior that we associate with nonlinear phenomena like asymmetry, time irreversibility, limit cycles, chaos, etc. For purposes of forecasting, an n-step ahead forecast is given by

$$\hat{X}_{t+n} = \sum_{i=1}^{s} \left(\phi_1^{(i)} X_{t-1} + \ldots + \phi_{m_1}^{(i)} X_{t-m_i} \right) I_{(r_{i-1}, r_i]}\left(X_{t+n-d} \right) \ , \tag{6}$$

where I_A is the indicator function of the set A.

Section 2 presents the concept of *value-at-risk* (*VaR*), specifically a VaR derived from a given SETAR model. Also, a review of a number of backtesting procedures is given. These procedures will be used to assess the accuracy of the proposed VaR

Section 3 introduces a methodology for fitting a SETAR model to the loss series generated from Philippine Stock Exchange index (PSEi) data from December 1, 2005 to October 10, 2012. From the fitted model, VaR estimates will be computed and assessed for accuracy using the various aforementioned backtesting procedures. VaR estimates from two other procedures will be compared in Section 4.

2. Value-at-Risk and Backtesting Procedures

In risk management, a popular measure of risk is the *value-at-risk* (oftentimes abbreviated as VaR.) VaR measures the largest potential loss one can incur in a given time horizon at a given confidence level. This simple definition consists of two components: (1) a time horizon Δ, usually Δ = 1 day or Δ = 10 days; and (2) a confidence level (1- α), typically 95% or 99%, that is, α is either 0.05 or 0.01. For this reason, VaR is often represented as $VaR_{1-\alpha}^{\Delta}$. If we represent "the loss at time *t*" by L_t and model it using (5), then this would suggest a *conditional VaR* given by the (1- α)100^th^ quantile of the distribution of L_t , i.e.,

$$VaR_{1-\alpha,t} = \inf\left\{y : F_{L_t}\left(y\right) \geq \left(1-\alpha\right)\right\}, \qquad (7)$$

where F_{L_t} is the distribution function of L_t .

Specifically, if the $\left\{\varepsilon_t^{(i)}\right\}$ innovations are independent Gaussian processes then conditional on ($L_{t+\Delta-1}$, $L_{t+\Delta-2}$,…, $L_{t+\Delta-m}$),

$$VaR_{1-\alpha,t}^{\Delta} = \sum_{i=1}^{s}\left[\phi_1^{(i)}L_{t+\Delta-1} + \ldots + \phi_{m_1}^{(i)}L_{t+\Delta-m_i} + \Phi^{-1}\left(1-\alpha\right)\sigma_{(i)}\right]I_{\left(r_{i-1},r_i\right]}\left(L_{t+\Delta-d}\right), \qquad (8)$$

where $\Phi^{-1}(1-\alpha)$ is the (1- α)100^th^ quantile of the standard normal distribution.

Backtesting procedures are often employed to assess the accuracy of VaR estimates. Although the conditional VaR given by (8) seems to be the most natural way to define the conditional quantile of the modeled loss at time t + Δ, it still needs to satisfy some basic requirements of agreement between the estimates and the data. For example, we expect that daily estimates for $VaR_{.99}$ to be exceeded about once every 100 days on the average. The backtesting procedure would then be a statistical test if the proportion of violations would not be significantly different from .01, at some given level of significance. Defining the indicator

$$\hat{I}_{t+1,\alpha} \equiv I\left(L_{t+1} > Va\hat{R}_{1-\alpha}^{t}\right)$$

and the statistic $\hat{\pi}_{\alpha} = \frac{\sum_{k=1}^{T} \hat{I}_{k,\alpha}}{T} = \frac{n*}{T}$, the **time of first failure (TUFF)** test is based on the number of observations before the first VaR violation. Under the null hypothesis H_0: $\alpha = \alpha^*$, the likelihood ratio test is

$$\text{LRTUFF} = 2\log\left[\frac{\left(\frac{1}{t*}\right)\left(1-\frac{1}{t*}\right)^{t*-1}}{\hat{\pi}_{\alpha*}\left(1-\hat{\pi}_{\alpha*}\right)^{t*-1}}\right] \overset{asymp}{\sim} \chi_1^2,$$

where t^* denotes time until the first exception.

Kupiec (1995) noted that the TUFF test has limited power to distinguish among alternative hypotheses because all observations after the first failure are ignored and proposed a nonparametric test based on the proportion of exceptions. Under the null hypothesis, the likelihood ratio test (also known as the **proportion of failure test** and also as the **unconditional coverage test**) is

$$LR_{POF} = LR_{UC} = 2\log\left[\frac{\left(1-\hat{\pi}_{\alpha*}\right)^{T-n*} \hat{\pi}_{\alpha*}^{n*}}{\left(1-\alpha*\right)^{T-n*} \alpha*^{n*}}\right] \overset{asymp}{\sim} \chi_1^2 .$$

Christoffersen (1983) argued that Kupiec's LR_{POF} test does not give any information about the temporal dependence of violations and ignores conditioning coverage, since violations could cluster over time. A **test for independence over time of VaR violations** uses a test statistic given by

$$LR_{IND} = 2\log\left[\frac{\left(1-\hat{\pi}_0\right)^{n_{00}} \hat{\pi}_0^{\,n_{01}} \left(1-\hat{\pi}_1\right)^{n_{10}} \hat{\pi}_1^{\,n_{11}}}{\left(1-\hat{\pi}_2\right)^{n_{00}+n_{10}} \hat{\pi}_2^{\,n_{01}+n_{11}}}\right] \overset{asymp}{\sim} \chi_1^2,$$

where $n_{00} \equiv$ number of two consecutive time periods with no violation,

$n_{10} \equiv$ number of time periods with no violation preceded by a time period with a violation,

$n_{01} \equiv$ number of time periods with a violation preceded by a time period with no violation,

$n_{11} \equiv$ number of two consecutive time periods of with violations,

$$\hat{\pi}_0 = n_{01} / (n_{01} + n_{00}),$$

$$\hat{\pi}_1 = n_{11} / (n_{11} + n_{10}),$$

$$\hat{\pi}_2 = (n_{01} + n_{11}) / (n_{01} + n_{00} + n_{11} + n_{10}).$$

In addition, Christoffersen (1998) also proposed **a test of correct conditional** coverage which tests the null hypothesis of an independent failure process with probability α^* against the alternative of a first order Markov failure process, given by

$$LR_{CC} = 2\log\left[\frac{\left(1-\hat{\pi}_0\right)^{n_{00}} \hat{\pi}_0^{\,n_{01}} \left(1-\hat{\pi}_1\right)^{n_{10}} \hat{\pi}_1^{\,n_{11}}}{\left(1-\alpha^*\right)^{T-n^*} \alpha^{*n^*}}\right] \overset{asymp}{\sim} \chi_2^2.$$

Haas (2001) proposed a test that incorporated that of Kupiec's and Christoffersen's. The test statistic given by

$$LR_{H\text{-}IND} = 2\sum_{i=1}^{v} \log\left\{\frac{\left(\frac{1}{d_i}\right)\left(1-\frac{1}{d_i}\right)^{d_i-1}}{\alpha^*\left(1-\alpha^*\right)^{d_i-1}}\right\},$$

where v = total number of VaR violations,

d_1 = time of first VaR violation,

d_i = time between the i^{th} and (i – 1)$^{\text{th}}$ VaR violation, i = 2,…, v has an asymptotic χ_v^2 distribution.

Similar to the LR_{CC}, combining $LR_{H\text{-}IND}$ and LR_{POF} gives the mixed Kupiec's test, i.e., $LR_{mix} = LR_{H\text{-}IND} + LR_{POF}$ which, under the null hypothesis H_0: $\alpha = \alpha^*$, has an asymptotic chi-square distribution with $v + 1$ degrees of freedom.

Finally, the proposed conditional VaR will also be subjected to the **dynamic conditional quantile (DQ) test** of Engle and Manganelli (2004) using the following statistic

$$DQ = \left(h_t' X_t \left[X_t' X_t\right]^{-1} X_t' h_t\right) / T\alpha^*\left(1-\alpha^*\right),$$

where $h_t = \hat{I}_{t,\alpha^*} - \alpha^*$ is the vector of demeaned violations,

X_t is a vector of instruments which might include lags of h_t,, VaR_t and its lags.

Under the null hypothesis that h_t and X_t are orthogonal, the proposed DQ statistic follows a χ_q^2 distribution with q = rank X_t. Using Monte-carlo experiments, Berkowitz et.al (2011) showed that the DQ test with $X_t = VaR_t$ appears to be the best backtest for 1% VaR models, and other backtests generally have much lower power against mispecified VaR models.

3. VaR Estimates Derived from a SETAR Model

Figure 1 shows the graphs of the values at the close of the trading day of the Philippine Stock Exchange index and its accompanying loss series for the period from December 1, 2005 to October 10, 2012. Losses were computed by taking the negative difference of log values and expressed in percentage, i.e.

$$Loss_t = -\log\left(\frac{PSEi_t}{PSEi_{t-1}}\right) \times 100\% .$$

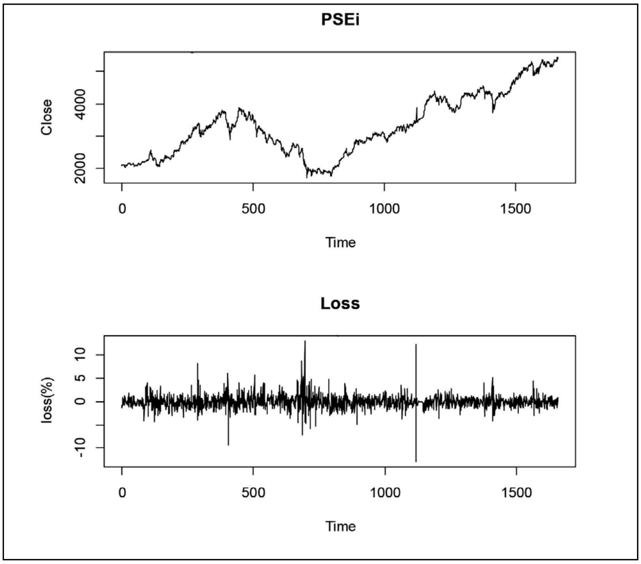

Figure 1. Time Plots of PSEi and Daily Percentage Loss from Dec. 1, 2005 to October 10, 2012.

Parameter estimation for the SETAR model is done via the `setar` () function included in the package `tsDyn` of R version 3.0.0. However, the said function needs additional parameters to be supplied by the user, particularly the embedding dimension, the number of regimes and the autogressive orders for each regime.

The embedding dimension is determined by the *minimum entropy heuristic* developed by Brandmaier (2012) and is implemented in R by the function `entropy.heuristic`() in the package `pdc`. Figure 2 shows that for the loss series, an embedding dimension of 6 minimizes the entropy.

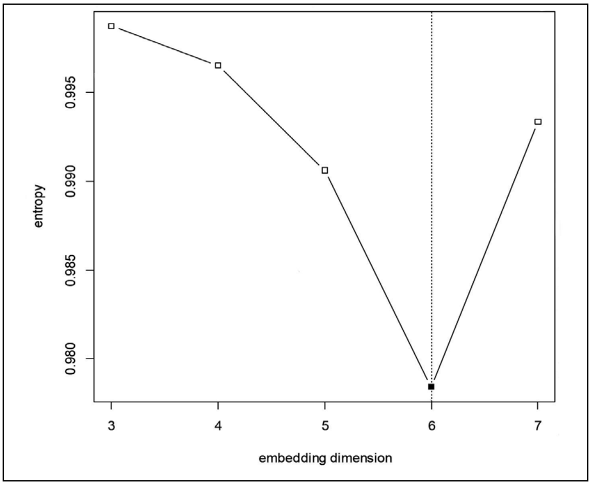

Figure 2. Plot of Entropy vs. Embedding Dimension

To determine the number of regimes, Hansen (1999) proposed the following. If we denote a SETAR model with m regimes by SETAR(m), then essentially SETAR(j) $\subset$ SETAR(k), for $j \leq k$. Simply put, the SETAR models are *nested*. The resulting LR test statistic for testing the SETAR(j) against SETAR(k) ($k > j$) is given by

$$F_{jk} = n\frac{SSE_j - SSE_k}{SSE_k} \ ,$$

where SSE_m is the sum of squared residuals in fitting a SETAR(m) model by *least squares* (*LS*) estimation. Table 1 presents the results Hansen's test, which favors more than 1 regime but that there is no significant difference between 2 and 3 regimes.

Table 1. Results of Hansen's test for linearity

	Test statistic	p-value
1 vs. 2	73.90587	0.000
1 vs. 3	87.51488	0.000
2 vs. 3	13.02591	0.592

Parameter estimation for different SETAR(2, d, p_1, p_2) models was done using the `setar()` function included in the `tsDyn` package, for $d, p_1, p_2 = 1, \ldots, 5$. The SETAR model with the smallest Akaike's Information Criterion (AIC) is chosen as the "best" model. For the given data, the "best"-fitting SETAR model is

$$L_t = \begin{cases} .06L_{t-2} \quad , & L_{t-5} \le 0.7193 \\ .13L_{t-1} - .13L_{t-2} - .09L_{t-3} - .1L_{t-4} - .126L_{t-5}, & L_{t-5} > 0.7193 \end{cases} \quad (9)$$

Figure 3 shows the VaR estimates computed using equations (8) and (9). Table 2 gives the results of the various backtesting procedures. For the DQ test, the suggestion by Berkowitz et al. (2011) to use the instrument $X_t = VaR_t$ was adopted. It can be seen that the proposed VaR fails mainly in Haas' independence test. As a consequence, this also leads to the failure of the mix test (its test statistic being the sum of the POF and H-IND tests.) These results are not completely surprising since the tests of independence tend to reject series if the VaR violations (i.e. when actual losses exceeds estimated VaR) are clustered. The same results were obtained by Nieppola (2009) for a number of equity, bond and options portfolios. This phenomenon has not completely escaped the notice of financial market analysts. In fact, such behavior has been given the term *volatility clustering* and is regarded as a stylized fact (Cont, 2001). The topmost panel in Figure 4 shows where the VaR violations for the SETAR model occur. Noteworthy is the presence not only of clustering but the maximum of such violations occurred during the latter months of the year 2008. These results, however, should not completely discount the VaR estimates since it should also be noted that VaR models are known to be accurate only under normal market conditions.

Table 2. Results of tests for VaR from SETAR

	POF	TUFF	IND	CC	H-IND	Mix	DQ
p-value	.548	.612	166	.019	.004	.003	.036
Result (at .01 level of significance)	Pass	Pass	Pass	Pass	Fail	Fail	Pass

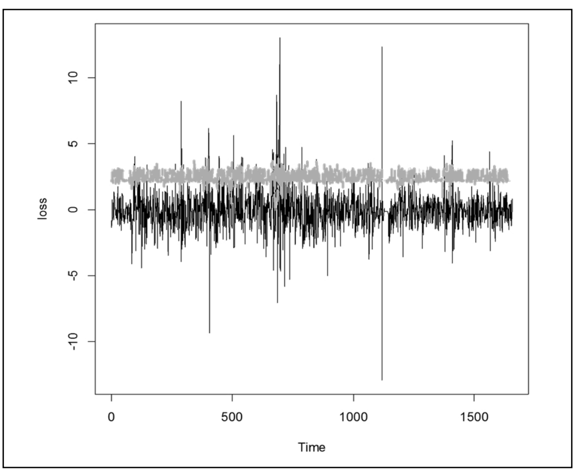

Figure 3. Time Plot of Conditional VaR

4. Comparison with Estimates from Historical Simulation and an APARCH model

It would be quite informative to compare the VaR estimates from the SETAR model with two other sets of estimates: (1) those derived from *historical simulation (HS)* and (2) those derived from another model, an AR(1)-APARCH(1,1) model. (The interested reader can refer to McNeil et al. (2005) for a discussion of historical simulation and to Ding et al. (1993) regarding the asymmetric power ARCH or APARCH model.)

Historical simulation is essentially an unconditional approach in which the VaR is just the quantile of the empirical distribution function of the loss values taken as a sample of independent observations. On the other hand, volatility clustering is popularly modeled by GARCH models (of which APARCH is a more generalized version). The resulting AR-APARCH model for the loss series is given by

$$Lt = \mu_t + \sigma_t^2$$

where $\mu_t = \underset{(.003)}{-0.0702} + \underset{(.01)}{.0739}L_{t-1}$ and

$$\sigma_t^2 = \underset{(.000)}{0.395} + \underset{(.000)}{0.170}\,(\,|\varepsilon_{t-1}| - \underset{(.000)}{0.359}\varepsilon_{t-1})^2 + \underset{(.000)}{0.619}\ \sigma_{t-1}^2$$

(Values inside parentheses are *p-values*).

Tables 3 and 4 show the results of the various backtesting procedures wherein both the HS and the AR-APARCH model pass Haas' independence test but fail in the more recent DQ test (at .01 level of significance.)

Table 3. Results of tests for VaR from Historical Simulation

	POF	TUFF	IND	CC	H-IND	Mix	DQ
p-value	.063	.470	.165	.000	.017	.020	.003
Result (at .01 level of significance)	Pass	Pass	Pass	Fail	Pass	Pass	Fail

Table 4. Results of test for VaR from APARCH

	POF	TUFF	IND	CC	H-IND	Mix	DQ
p-value	.897	.461	.851	.705	.153	.158	.003
Result (at .01 level of significance)	Pass	Pass	Pass	Pass	Pass	Pass	Fail

Figure 4 shows almost similar results for all three sets of estimates, particularly the presence of clustering at various epochs of the time interval, most especially during the latter part of 2008. SETAR VaR estimates have the smallest number of VaR violations, followed by the AR-APARCH VaR estimates. If we define *mean shortfall* as the average amount of excess loss over VaR estimates where VaR violations occur then the mean shortfall for the SETAR VaR lies between that of HS and the AR-APARCH VaR. (Care should be taken not to confuse mean shortfall with *expected shortfall*, although the former can serve as an estimate of the latter). Although the formal tests did not reject the Haas'-independence for the HS and AR-APARCH VaR violations, the similarities in the locations and magnitude of these violations can clearly be seen by comparing the graphs in Figure 4.

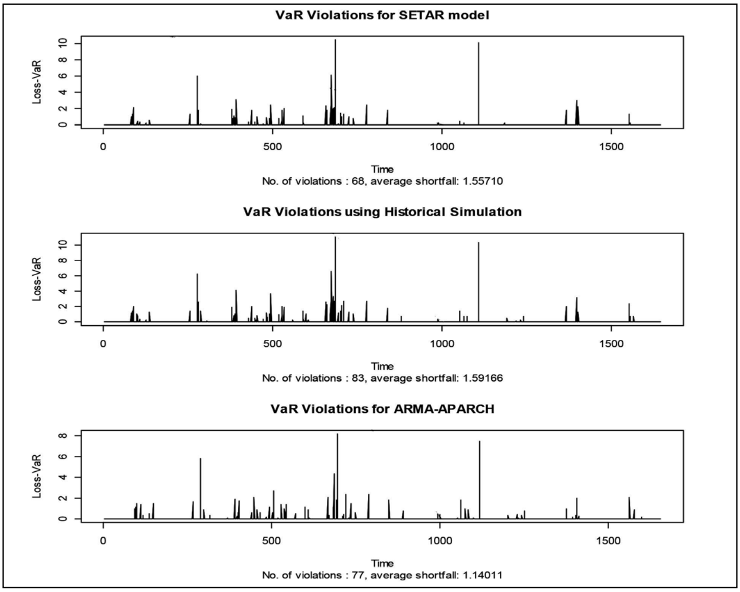

Figure 4. VaR violation for the SETAR model, Historical Simulation and AR-APARCH model

5. Conclusion

Despite the shortcomings of the VaR estimates derived from it, the SETAR remains a viable model for fitting time series data for the following reasons:

1. A number of nonlinear phenomena can be exhibited by SETAR models. For example, one stylized fact is volume/volatility correlation, defined by Cont (2001) as "trading volume is correlated with all measures of volatility." In the dynamical systems literature, this translates to the concept known as "amplitude-frequency dependence." Other nonlinear phenomena, like asymmetry and time-irreversibility (Tong, 1990) can be modeled as SETAR.

2. Methodologies and computer software are available for fitting, testing and making inferences about SETAR models.

3. Ease of interpretation. Unlike the variants of GARCH models, SETAR models offer the same ease of interpretation offered by autogressive (AR) models. That is, in each regime, the behavior of the series is nothing more than a linear combination of previous values. The presence of thresholds and regimes can bring additional insight into the behavior of the time series.

REFERENCES

BERKOWITZ, J., CHRISTOFFERSEN, P., PELLETIER, D., 2011, Evaluating value-at-risk models with desk-level data, *Management Science* 57:2213-2227.

BRANDMAIER, A.M., 201, Permutation distribution clustering and structural equation model trees. Dissertation. Saarland University. Saarbruec. Available at: (URL:http://scidok.sulb.uni-saarland.de/volltexte/2012/4545/pdf/Andreas_M_Brandmaier_Dissertation.pdf)

CAMPBELL, S. D., 2005, A review of backtesting and backtesting procedures. Available at: (URL: http://www.federalreserve.gov/pubs/feds/2005/200521/200521pap.pdf)

CONT, R., 2001, Empirical properties of asset returns: stylized facts and statistical issues, *Quantitative Finance* 1: 223-236.

DI NARZO, A.F., AZNARTE, J.L. and STIGLER, M., 2009, tsDyn: Time series analysis based on dynamical systems theory. R package version 0.7. Available at: (URL: http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/tsDyn.pdf}

DING, Z., GRANGER, CWJ., ENGLE,. R.F., 1993, A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1: 83-106.

ENGLE, R.F. AND MANGANELLI, S., 2004, CAViaR: conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics* 22 (4): 367-381.

KUPIEC, P.H. 1995, Techniques for verifying the accuracy of risk measurement models, *The Journal of Derivatives* 3: 73-84.

HAAS, M., 2001, *New Methods in Backtesting*, Financial Engineering Research Center Caesar, Bonn.

HANSEN, BRUCE E., 1999, Testing for Linearity, *Journal of Economic Surveys* 13(5): 551–576.

MCNEIL, A., FREY, R. & EMBRECHTS, P., 2005, *Quantitative Risk Management*, Princeton University Press, New Jersey.

NIEPPOLA, O., 2009, Backtesting value-at-risk models. Unpublished Thesis. Helsinki School of Economics. Available at:(URL:https://aaltodoc.aalto.fi/bitstream/handle/123456789/181/hse_ethesis_12049.pdf?sequence=1)

STIGLER, M., 2010, tsDyn: Threshold cointegration: overview and implementation in R R package version 0.7-2. Available at: (URL: http://stat.ethz.ch/CRAN/web/packages/tsDyn/vignettes/ThCointOverview.pdf)

TONG, H., 1990, *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

VIRDI, N.K., 2011, A Review of Backtesting Methods for Evaluating Value-at-Risk. *Intl Review of Business Research Papers* 8(4): 14-24.