Technical Report: Nonlinearity in the Univariate China Toward Taiwan Net Count Series *

John Freeman

July 9, 2012

...the world is more nonlinear than we think...
Nassim Nicholas Taleb, *The Black Swan* 2007: 88.

This report briefly reviews some simple nonlinear time series models. It then fits a univariate model of this type to a series for directed dyadic behavior in the Taiwan straits, specifically, the monthly net of materially cooperative and materially conflictual events directed by China toward Taiwan in the period 1998:1-2011:3. The analysis illustrates the problem one encounters in (frequentist) pretesting for nonlinearity and also highlights the risks of overfitting the data–the ability to find a good fitting threshold autoregressive model when the pretests for the data are inconclusive. More generally, the results provide a methodological benchmark for the fuller, multiequation, nonlinear modeling efforts in Lin et al. (in progress) and in Brandt et al (2012).[1]

1 Nonlinear Time Series Models–A Brief Review

Nonlinear time series models describe processes which exhibit asymmetries and(or) sudden bursts in amplitude at irregular intervals. Nonlinear models also are useful for analyzing time series processes that are characterized by time irreversibility (Tong 1990: Section 1.5). The conditions for stochastic stability for nonlinear models of the form $y_t = f(y_{t-1}, \epsilon_t)$ are summarized by Granger and Teräsvirta (1993: 12; their original source is Lasota and Mackey 1989). A more general treatment of stochastic stability and of stationary (densities) distributions for nonlinear time series models is Tong (1990).[2]

*This work is supported by the NSF under grant SES 0921018. The comments of Patrick Brandt and the research assistance of Geoff Sheagley and Andrew Lucius are gratefully acknowledged. Please do not quote without the author's permission

[1]The discussion in this report follows that in the last part of Freeman and Jackson (2012). The application follows the steps in a univariate analysis of Green's macro political partisanship series in Freeman (2012). Both sources explain the connection between these simple nonlinear models and the concepts associated with path dependence.

[2]There are a variety of nonlinear time series models including generalized autoregressive (GAR), the bilinear, and multiple forms of threshold autoregressive models such as STAR, LSTAR, and ESTAR. Still another is the markov switching time series model. For an introduction to these models see such works as Enders (2010: Chapter 7) and Granger and Teräsvirta (1993). The latter source cites Quinn (1982) for the

1.1 Univariate Nonlinear Time Series Models

The threshold autoregressive model, TAR, or self-excited threshold, SETAR, model, is one of the most widely used nonlinear time series models. It has been has been employed to model a variety of physical and social processes. Tong (1990) argues that thresholds are generic concepts. He shows how the SETAR model can be used to model sunspot and animal (lynx) population series. Enders (2010) reviews TAR models of unemployment, models that capture the fact that when an economy is in a recession and unemployment is above a threshold the speed of recovery (job growth) might be slow whereas if unemployment is below a threshold unemployment might gravitate towards its long term equilibrium much more rapidly. The Simple TAR model is:

$$y_t = \begin{cases} a_1 y_{t-1} + \epsilon_t & \text{if } y_{t-1} > r \\ a_2 y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq r \end{cases}$$

where r is the threshold. This data generating process is a combination of two simple AR(1) processes. Which AR(1) process occurs depends on whether the previous value, y_{t-1} is above or below its threshold, r. The behavior of the Simple TAR model will differ depending on which of the two regimes apply. It can be shown that this model is geometrically ergodic if $a_1 < 1, a_2 < 1$ and $a_1 a_2 < 1$ (Tong 1990: 130-1).

A slightly more complicated version of the TAR model allows for each AR processes to have different constants and different errors terms. The expected values of the two AR process then are distinct as are the variances of the errors. This could be called a Basic TAR model.[3] An example of such a model is:

$$y_t = \begin{cases} a_{10} + a_1 y_{t-1} + \epsilon_{1t} & \text{if } y_{t-1} > r \\ a_{20} + a_2 y_{t-1} + \epsilon_{2t} & \text{if } y_{t-1} \leq r \end{cases}$$

Under certain conditions, each AR process has a different expected value, either $\frac{a_{10}}{1-a_1}$ or $\frac{a_{20}}{1-a_2}$. So its limiting behavior will switch between adjustment to two different long term values. The conditions for geometric ergodicity of such models have been derived by Chan et al (1985); one such condition is $a_1 < 1, a_2 < 1, a_1 a_2 < 1$. Consider, for purposes of illustration, the following model

$$y_t = \begin{cases} 1.5 - 0.9 y_{t-1} + \epsilon_t & \text{if } y_{t-1} > 0 \\ -0.4 - 0.6 y_{t-1} + \epsilon_t & \text{if } y_{t-1} \leq 0 \end{cases}$$

Tong (1990: Section 4.2.4.3) shows how a numerical method can be used to estimate the stationary density of this particular process. This density is depicted in Figure 1 above.

Some of these models allow for nonstationary behavior. One of the most simple is the

conditions for the stability on the bilinear model. But it also notes that stability results are not always available. A more general review of nonlinear models is Tong (1990). Here I focus on a few simple examples of such time series models.

[3] My nomenclature differs somewhat from Enders (2010: 439) who uses the word "Basic" to describe a TAR model with no constants. I called this the "Simple TAR model" above. In addition, in his Introduction, Enders (2010: 429-430) uses the idea of a single, long term "attractor" for a nonlinear TAR process.

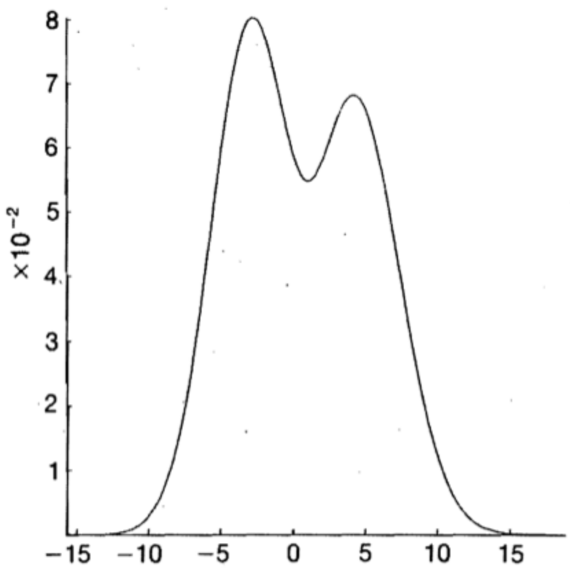

Figure 1: Stationary Density of Illustrative Basic TAR Model. Source Tong (1990: Section 4.2.4.3)

Equilibrium-TAR model:

$$x_t = \begin{cases} x_{t-1} + \mu_t & \text{if } |x_{t-1}| < k \\ \rho x_{t-1} + \mu_t & \text{otherwise} \end{cases}$$

where ρ is a constant, $|\rho| < 1$, and $\mu_t \sim N(0, \sigma_\mu^2$.

A somewhat more complex model of this kind is the Band-TAR model. Enders (2010: 446) provides an illustration. Let $s_t = r_{Lt} - r_{St}$ be the spread between long and short term interest rates. Assume this spread follows a simple AR(1) process with constant coefficients, more specifically,

$$s_t = a_0 + a_1 s_{t-1} + \epsilon_t \tag{1}$$

where ϵ_t is the familiar white noise error process. Assume further than the AR(1) process is covariance stationary hence its expected value is $\frac{a_0}{1-a_1}$. Call this long-run value $\bar{s}$. This allows us to rewrite (1) as an adjustment process of the form

$$s_t = \bar{s} + a_1(s_{t-1} - \bar{s}) + \epsilon_t. \tag{2}$$

where, again, ϵ_t is a white noise error term. Then the Band-TAR model can be expressed in the form

$$s_t = \begin{cases} \bar{s} + a_1(s_{t-1} - \bar{s}) + \epsilon_t & \text{if } s_{t-1} > \bar{s} + c \\ s_{t-1} + \epsilon_t & \text{if } \bar{s} - c < s_{t-1} \leq \bar{s} + c \\ \bar{s} + a_2(s_{t-1} - \bar{s}) + \epsilon_t & \text{if } s_{t-1} \leq \bar{s} - c \end{cases}$$

Several points should be made about these models. First, conceptually, they are associated with the idea of transaction cost or arbitrage boundaries. Agents supposedly monitor the process and decide that once the variable exceeds certain values, the (net) benefit of intervening (incurring a transaction cost) to drive it back into the intermediate (locally non-stationary) range exceeds the cost of (foregoing the intervention) and allowing the process to be (globally) non-mean reverting (Balke and Fomby 1997). Second, analysis shows that the statistical power of conventional tests for nonstationarity depend on the parameters of these models. For instance, Pippinger and Goering (1993) demonstrated how the power of the Dickey Fuller test to detect mean revision in the Equilibrium TAR model depends on ρ, k, and σ^2_μ. The wider the interval, k, for instance, the more time the process spends in the nonstationary region. Hence, even if ρ is small, the Dickey Fuller test has low power. In this context, Pippinger and Goering conceive of "equilibrium" as the *continuum* of values in the interval [-k,k] (*Ibid.*, fn. 4). So, within this range, the process is equilibrium (path) dependent. Globally, however, under certain conditions, such a process actually is stationary. What is required for global stationarity is that the process be mean-reverting in the "outer regimes." This can occur even if these regimes the random walks with drifts as long as the drift parameters "act to push the series back to the equilibrium band" (Balke and Fomby 1997: 630). Once again, locally, within this band the process is equilibrium dependent while, globally, it is stationary.[4]

1.2 Multivariate, Nonlinear Time Series Models

Jackson and Kollman (2010, 2012) analyze strongly restricted, nonlinear, multivariate time series regression models in which one variable is posited to be exogenous. They show how such models can exhibit path and near-path dependence and, concomitantly, equilibrium dependence. We refer interested readers to their papers.

As regards weakly restricted models, the idea of "threshold cointegration" addresses the possibility that two or more series are nonstationary but share a common trend(s). Enders desribes a model of this kind.[5] Let r_{LT}, r_{St} represent the interest rate on ten year government securities and the federal fund rate, respectively. Assume each series is I(1) and that they are cointegrated. The model captures regime shifts in terms of how changes in the interest rate spread, $s_t = r_{Lt} - r_{St}$, increasing vs. decreasing, translate into different rates of error correction. In this case, there is no error correction when $s_{t-1} = \beta$. This is a threshold model of the momentum type, M-TAR:

$$\begin{aligned} \Delta r_{Lt} &= \alpha_{11} I_t[s_{t-1} - \beta] + \alpha_{12}(1 - I_t)[s_{t-1} - \beta] + A_{11}(L)\Delta r_{L,t-1} + A_{12}\Delta r_{S,t-1} + \epsilon_{1t} \\ \Delta r_{St} &= \alpha_{21} I_t[s_{t-1} - \beta] + \alpha_{22}(1 - I_t)[s_{t-1} - \beta] + A_{21}(L)\Delta r_{L,t-1} + A_{22}\Delta r_{S,t-1} + \epsilon_{2t} \end{aligned}$$

where the α terms are adjustment coefficients, $s_t = r_{Lt} - r_{St}$, the $[s_{t-1} - \beta]$ terms are cointegrating vectors, the $A(L)$ terms are lag operators, and the I_t variable is an indicator function defined as

[4]To illustrate this point, Balke and Fomby analyze the Returning Drift (RD) Threshold Model:

$$z_t = \begin{cases} -\mu + z_{t-1} + \epsilon_t & \text{if } z_{t-1} > \theta \\ z_{t-1} + \epsilon_t & \text{if } |z_{t-1}| \leq \theta \\ \mu + z_{t-1} + \epsilon_t & \text{if } z_{t-1} < \theta \end{cases}$$

where μ is the drift parameter and the ϵ_t are mean zero random disturbances.

[5]This example is a simplified version of an example in Enders (2010: 481).

$$I_t = \begin{cases} 1 & \text{if } \Delta s_{t-1} > 0 \\ 0 & \text{if } \Delta s_{t-1} \le 0 \end{cases}$$

For this model then, the rate of adjustment to the moving equilibrium between the two phat outcome dependent processes varies depending on whether in the previous period s_t was increasing or decreasing.

Balke and Fomby (1997) is a more general treatment of threshold cointegration. They introduce the idea of a discontinuous adjustment to long-run equilibrium, a process that adjusts to long-run equlibrium at some times but not others. Again, the motivation assumes there are agents (policy makers) that sometimes find it in their interest to force to variables to trend together while in other cases they allow two processes to diverge from long term equilibrium. They explore, in the spirit of the Engle and Granger approach, in a Monte Carlo investigation, the power and size properties of five different tests for cointegration for the Equilibrium-TAR and Band-TAR models described above and the RD-TAR model described in fn. 4. They conclude standard linear methods for testing for cointegration work well in the presence of threshold cointegration. Balke and Fomby then proceed to develop a method to detect two threshold cointegration based on the concept of arranged autoregression.[6]

2 Illustration: The Dynamics of China's Behavior Toward Taiwan

2.1 Univariate Analyses

Our series for Chinese behavior toward Taiwan is the number of directed monthly events of a materially cooperative nature minus the number of directed monthly events of a materially

[6]So in the Balke and Fomby (1997) the models are written in terms of the *error term from the cointegrating regression*. Sometimes this error is stationary connoting long-term equilibration of the two integrated series (cointegration), and sometimes the error is nonstationary connoting a lack of long-term equilibration (and absence of cointegration). Their simple example is the model:

$$y_t + \alpha x_t = z_t, \qquad \textit{where} \quad z_t = \rho^{(i)} z_{t-1} + \epsilon_t \tag{3}$$

$$y_t + \beta x_t = B_t, \qquad \textit{where} \quad B_t = B_{t-1} + \eta_t \tag{4}$$

where the ϵ_t, η_t are white noise disturbance terms. Then the value of ρ varies depending on the magnitude of z_t:

$$\rho^{(i)} = \begin{cases} 1 & \text{if } |z_{t-1}| \le \theta \\ \rho, \quad |\rho| < 1 & \text{if } |z_{t-1}| > \theta \end{cases}$$

When the absolute value of the first lag in the error, z_{t-1} is less than the threshold, θ, $\rho^{(i)} = 1$, and the two I(1) variables, x_t, y_t, do not revert to a long-run equilibrium. But if the first lag of this same error, is greater than θ in absolute value , $\rho^(i) = \rho$ and $|\rho| < 1$ so the two variables do move towards some equilibrium. Balke and Fomby proceed to present the most general version of this model and then study in their Monte Carlo analyses of cointegration tests, the Equilibrium TAR, Band-TAR, and RD-TAR versions of the above model.

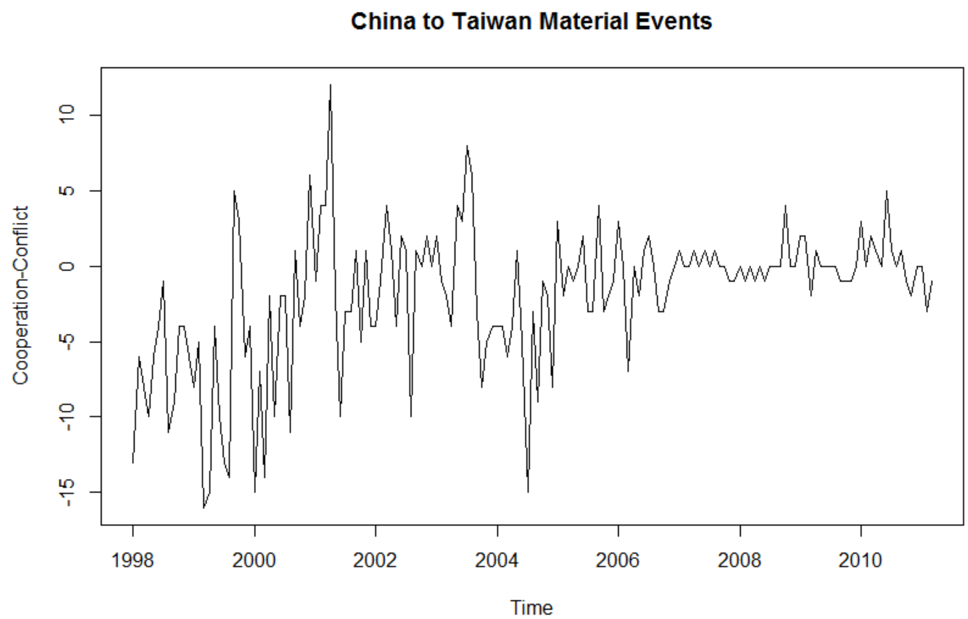

Figure 2: Monthly Materially Cooperative Events Minus Materially Conflictual Events Directed by China towards Taiwan, $C2TM_t$, 1998:1-2011:3.

conflictual nature as coded by the Event Data Project at Pennsylvania State University.[7] We denote it by $C2TM_t$; it is depicted in Figure 2. Visually, the series exhibits more variance in early than in more recent months. Whether it is composed of multiple regimes the switches between which depend on particular behavioral thresholds is not clear from an occular examination. The series appears to be stationary. In fact, simple Dickey Fuller tests (Table 1) indicate the series is stationary. Although the ϕ statistics leave open the possibility of a deterministic trend, such a trend is not evident in the series (Table 1). However, as noted aboved, these statistics are problematic. The Dickey-Fuller unit root test assumes linearity. And, under certain conditions, it can lack power if the underlying process is nonlinear (Enders 2010: Section 11; Enders and Granger 1998: Section 1, Pippinger and Goering 1993.).[8]

We begin by pretesting for evidence of nonlinearity in $C2TM_t$. Investigations of this kind usually include descriptive analysis of the series, for instance, study of the histogram and autopairs plots for the series (Tong 1990: esp. pps. 362-375). In the interest of brevity we relegate these analysis to the Appendix. Simply put, these descriptive analyses show little evidence of nonlinearity. Rather, we follow Enders (2010: Chapter 7, Section 3) in

[7]For the period we analyze this amounts to about 480,000 events. The CAMEO coding format was employed (Gerner et al. 2009). CAMEO events categories 06 to 09 are classified as material cooperation while categories 15 to 20 are classified as material conflict. Examples of each type of event are sending aid and destruction of property, respectively. For a fuller description of the series see Brandt et al. 2012.

[8]The KPSS statistics for $C2TM_t$ were calculated with RATS. RATS used the Schwert criterion to set the maximum lag at 13; RATS weighted autocovariances by the Bartlett kernel. Briefly, the KPSS statistic was statistically significant at all 13 lags at the 10% level and at lags 0-7 at the 5% level.

Specification	Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Zero lags in ADF				
No Intercept or Trend				
τ_1	-7.52	-2.58	-1.95	-1.62
Intercept and No Trend				
τ_2	-8.28	-3.46	-2.88	-2.57
ϕ_1	34.33	6.52	4.63	3.81
Intercept and Trend				
τ_3	-9.55	-3.99	-3.43	-3.13
ϕ_2	30.45	6.22	4.75	4.07
ϕ_3	45.62	8.43	6.49	5.47
Four lags in ADF				
No Intercept or Trend				
τ_1	-3.09	-2.58	-1.95	-1.62
Intercept and No Trend				
τ_2	-3.40	-3.46	-2.88	-2.57
ϕ_1	5.83	6.52	4.63	3.81
Intercept and Trend				
τ_3	-4.06	-3.99	-3.43	-3.13
ϕ_2	5.60	6.22	4.75	4.07
ϕ_3	8.34	8.43	6.49	5.47

Table 1: Results for Dickey Fuller Tests on $C2TM_t$. R output

Coefficients and Statistic	R Estimates	STATA Estimates
AR(1)	0.38(.07)	0.38(.06)
Seasonal AR(7)	0.32(.08)	0.32(.07)
Constant	-2.10(0.75)	-2.10(.80)
σ^2	16.21	16.24
AIC	903.05	903.05
$Q(\chi^2(df))$	34.66(34)	40.7(40)
	p = .436	p=.441

Table 2: Best Fitting Linear Model for $C2TM_t$: AR(1)-Seasonal AR(7) with a constant. Numbers in parentheses are standard errors. All the coefficients for seasonal coefficients other than AR(7) were set to zero. These restrictions are incorporated in degrees of freedom for Q statistic.

emphasizing the results of several portmanteau tests for independence, the results of which include, implicitly or explicitly, the possibility of nonlinearity: the McCleod-Li, RESET, and BDS (delta) tests. We also implemented a test for linearity vs. a specific TAR model (a supremum F test). In these analyses we used a combination of STATA, RATS, and R code.[9]

The first step in implementing the portmanteau tests is to estimate a linear model for the $C2TM_t$ series. The pacf for the series has a statistically significant spike at the seventh lag. This is surprising since there is no obvious substantive reason to expect seasonality in this (net) monthly material events series over the 14 year time period. Numerous linear models were estimated. The best fitting model was a AR(1)-Seasonal AR(7) model with a constant. Table 2 reports the estimates from the R and RATS packages [The collection of linear models which we fit are reported in the Appendix].

Two of the portmanteau tests have null hypotheses of linearity. The Regression Error Specification Test (RESET) regresses the residuals from the best fitting linear model on the regressors used in the estimating equation and *powers* of the fitted values from this equation. An F test is used to assess the joint statistical significance of the coefficients on the powers of the fitted variables.[10] For a test in which the fitted values were raised to the second and third powers, R returned a F statistic of .0006 (3, 154). This statistic has a p value of .999. This result means we cannot reject the null of linearity.

[9]The RATS programs are provided by Enders in his Instructor's Resource Guide. We employed the time series programs in the core R package as well as another package, tsDyn, version 0.7. This version of tsDyn appeared in 2008. Apparently it has not been updated.

[10]As Enders (2010: 436) explains, one first estimates a linear model and obtains the fitted values of the variable of interest, say, $\hat{y}_t$. Then the following equation is estimated:

$$e_t = \delta z_t + \sum_{h=2}^{H} \alpha_h \hat{y}_t^h \qquad for \qquad H \geq 2 \tag{5}$$

where e_t represents the estimated residuals, z_t is the vector of explanatory variables in the ARIMA model including the constant. Again, the RESET, distributed F, assesses the joint statistical significance of the α_h's.

The McLeod-Li test is the same as that used to detect ARCH type errors. In this case, one analyzes the sample autocorrelation coefficients between the *squared* residuals of the best fitting linear model. A Ljung-Box Q statistic is calculated for these squared autocorrelation coefficients. The statistic again has a χ^2 distribution. Figure 3 reports the results of the McLeod-Li test for the residuals from our AR(1)-seasonal AR(7) model of $C2TM_t$. The null of linearity is rejected at all but the first lag.

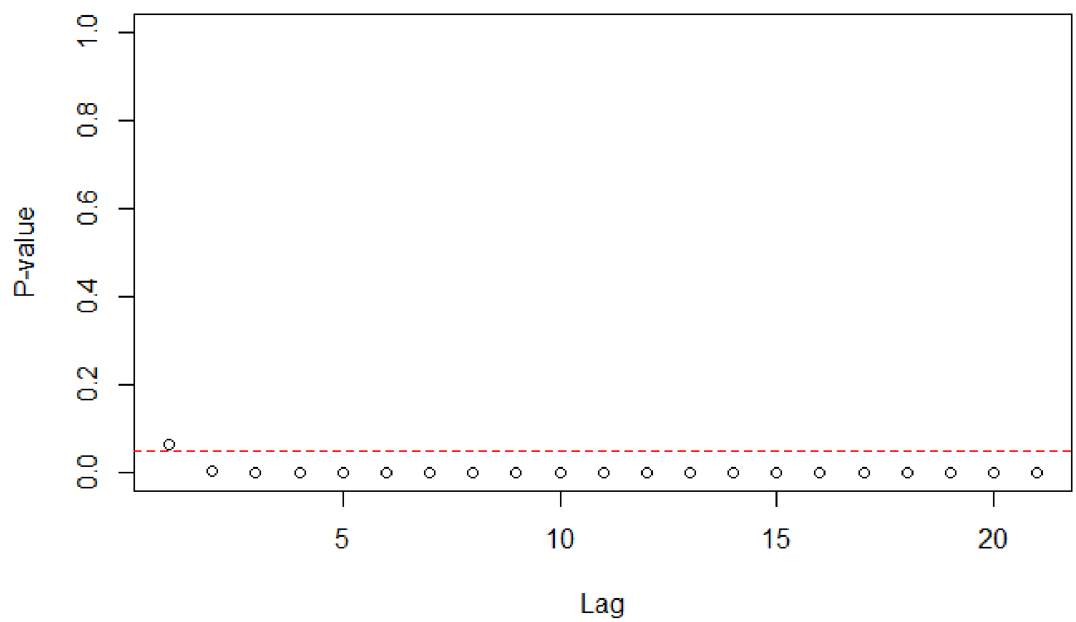

Figure 3: McLeod-Li Test P values for Different Lags for Squared Values of Residuals from our AR(1)-Seasonal AR Model of Chinese Directed Material Behavior Toward Taiwan

The BDS test is a third portmanteau test. It is named after Brock, Dechert, Scheinkman, and LaBaron (1996). The BDS test analyzes the "spatial dependence" of a time series–how "close" pairs of observations m lags apart are conditional on intermediate values of the series. The distance metric is set (by the tsDyn package) to four values: .5, 1, 1.5 and 2 times the standard deviation of the series. m is called the embedding dimension.[11]

[11]This terminology, as described in tsDyn Version 7 (p. 5) is based on the following formal representation of a discrete time univariate stochastic process, $[X_t]_{t \in T}$. The "map" for this process is written:

$$X_{t+s} = f(X_t, X_{t-d}, \ldots, X_{t-(m-1)d}; \theta) + \epsilon_{t+s} \tag{6}$$

where $[\epsilon_t]_{t \in T}$ is white noise and also independent of X_{t+s}, and f is a generic function from R^M to R. These models are abbreviate NLAR(m) which denotes Nonlinear AutoRegressive models of order m. The parameters m, d, s and θ are the embedding dimension, time delay, and forecasting steps, and coefficients on the lag terms of the model, respectively.

There are BDS (delta)tests for independence and linearity. Diks and Manzan (2002: 3) apply these tests to the original (level) time series. They argue that "the advantage over testing for dependence in residuals is that [by using the raw data] the lag dependence in the time series is preserved"[12] However, Granger and Teräsvirta (1993: 91) say the test should be applied to the residuals from the best fitting linear model. If the null hypothesis is rejected "one can conclude that nonlinearity is present but its form is not determined. It can be chaos or a nonlinear stochastic process." Enders notes that rejection of the null based on the BDS test indicates various types of misspecification including but not necessarily implying nonlinearity (Enders 2010: 437).[13] The small sample performance of the BDS test is not good. Bootstrapped confidence intervals are recommended (*Ibid.*: 91, 102; Enders 2010: 437). The delta test produces bootstrapped based, one sided p-values (Manzan 2003; Diks and Manzan 2002). Delta tests for independence and for linearity were performed on both the raw $C2TM_t$ data and on the residuals from our best fitting linear model.[14]

Briefly, the results are mixed. The p-values for some distances (ϵ) indicate rejection of the null of independence for the (raw) level series for $C2TM_t$ and rejection of the null of independence for the residuals from our linear model for this series. The results for the test of linearity of the raw data are less definitive. There is less support for the alternative of nonlinearity in this case.

Enders (2010: 449-450) describes one additional test. This is the test of the null of a simple linear model against an alternative, nonlinear SETAR model of the same structure. It uses Hansen's supremum F test, the critical values for which are determined by a bootstrapping procedure. We implemented this test for our AR(8) model of $C2TM_t$. The result suggest nonlinearity. The RATS program indicate a maximum F value for the TAR(8) model

[12]Diks and Manzan (2002) develop information theoretic tests for independence and linearity based on the idea of conditional mutual information (intermediate lag values of the series).

[13] Granger and Teräsvirta (1993: 36) explain how the test has power against white noise chaotic processes as well as against a variety of nonlinear stochastic processes. They (1993: 90-91) provide a full description of the BDS test and its value in testing for chaotic dynamics. What follows is a condensation of their description. Let $X_{t,m}$ denote a set of consecutive terms from a series x_t such that $X_{t,m} = (x_t, x_{t+1}, \ldots x_{t+m-1})$. A pair of vectors, $X_{t,m}$ and $X_{s,m}$, are said to be ϵ apart if the following relationship holds for each of pair of the corresponding terms:

$$|x_{t+j} - x_{s+j}| \leq \epsilon, \qquad j = 0, 1, \ldots, m-1. \tag{7}$$

The correlation integral, $C_m(\epsilon)$ is the limit of T^{-2} times the number of pairs (s,t) that are close in the sense (see *Ibid.* Section 3.3.1). Then the BDS statistic is:

$$S(m, \epsilon) = \hat{C_m}(\epsilon) - [\hat{C_1}(\epsilon)]^m \tag{8}$$

for some choice of m and ϵ Under the null that x_t is i.i.d. $\sqrt{TS(m, \epsilon)}$ has a normal distribution with mean zero and a variance that is a function of m and ϵ.

[14]For reasons that are not clear, the tsDyn program returned an error message for the linearity tests for the residuals from our linear model. Note that the tsDyn R package describes the delta test routine as experimental. The illustrations in the package are for delta independence and linearity tests for raw data not residuals from linear models. The illustration in the package reports rejection of the null for independence of the well known lynx data but *not* for the null of linearity of these data (*Ibid.* p. 16). At this point, the authors of tsDyn admit the results are anomalous and stress the delta test routines are experimental. Again, we can find no updated version of tsDyn since 2008.

Epsilon	2.343	4.686	7.030	9.737
Independence				
m=2	0.02	0.02	0.02	0.02
m=3	0.10	0.02	0.02	0.02
Linearity				
m=2	0.16	0.06	0.08	0.36
m=3	0.14	0.02	0.02	0.04

Table 3: Delta Test Results for Independence and for Nonlinearity of Chinese Material Directed Behavior Toward Taiwan. Raw data. Epsilon is distance as calculated by formula in the text. M again is the embedding dimension for the intermediate values of the series. Values are one sided p values generated by bootstrapping.

Epsilon	2.019	4.038	6.057	8.076
m=2	0.02	0.02	0.02	0.02
m=3	0.18	0.06	0.02	0.02

Table 4: Delta Test Results for Independence of Residuals from Linear Model of Chinese Material Directed Behavior Toward Taiwan. Entries are one sided p values generated by bootstrapping.

of 18.7523 at a threshold of -5 (Figure 6). The bootstrap p value based on 5000 replications for this F statistic is $< .0001$.[15]

Of course, there is no reason to assume that the SETAR model is of the AR(8) form. Therefore, using the selectSETAR routine in tsDyn we evaluated a variety of possible TAR models, 1323 to be exact.[16] On the basis of the routine's pooled-AIC criterion the best fitting model for $C2TM_t$ had a threshold of -1, threshold delay of 2, and seven lags in both the high and low regimes. This model then was estimated. The result was:

$$C2TM_{t+1} = \begin{cases} .27C2TM_{t-1} - .39C2TM_{t-4} + .29C2TM_{t-6} & \text{if } C2TM_{t-2} > -1 \\ .32C2TM_{t-1} + .24C2TM_{t-4} - .34C2TM_{t-6} + .35C2TM_{t-7} & \text{if } C2TM_{t-2} \leq -1 \end{cases}$$

where the statistically insignificant coefficients ($p > .05$) are not reported. The fit statistics for this model are: AIC 418, and MAPE 106%. Figure 7 is its regime switching plot. 52.6 % and 47.37 % of the observations, respectively, are in the low and high regimes.[17] As discussed in section 3.2 above, each regime exhibits a different kind of sequence outcome

[15]We could not get the RATS program to estimate a threshold model for the constant+AR(1)-seasonalAR(7 model.

[16]We set the forecast steps and regular delay parameters for the model to the defaults of 1. We then explored the fit of 3 threshold delays (1,2,3), alternative (independent) lag structures for the high and low regimes (each ranging between 1 and 7 in each regime), and 9 possible thresholds corresponding the the $C2TM_t$ levels remaining after trimming the lowest and highest 15% of the values(Figure 5). This produced 3x7x7x84=1323 models.

[17]The first seven observations appear to be treated as initial conditions and hence are not reported by tsDyn on the plot. Geoff Sheagley produced Figure 7 which includes these observations.

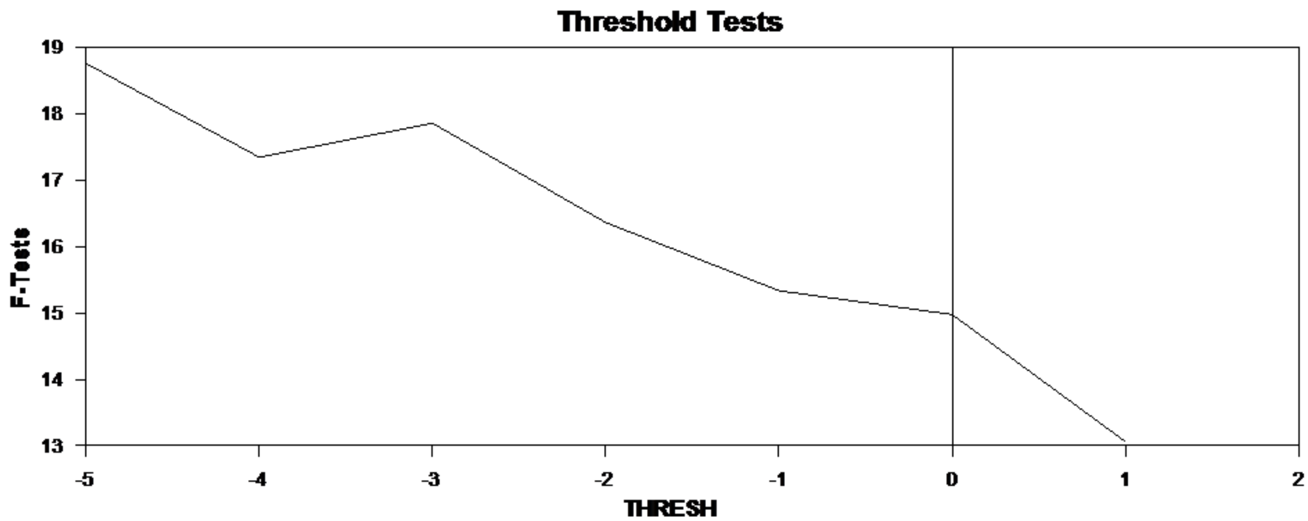

Figure 4: Maximum F Statistic for AR(8) Model of Net Chinese Material Behavior Toward Taiwan

dependence; the limiting behavior of each alone approaches a different expected value. Like Figure 2, globally, the stationary density of this nonlinear macropartisanship process has two humps not a single peak.

3 Conclusion

This note illustrates the problems associated with (frequentist) pretesting and the possible risks of overfitting data. We found a substantively inexplicable seasonal structure in our series. The results of two well known portmanteau tests were inconsistent. The Delta test results also were inconsistent. Enders version of Hansen's supremum F test for nonlinearity in a model with a single variance indicated nonlinearity in $C2TM_t$. In view of these results it is not clear how much stock we should put in the fact that tsDyn was able to find a plausible best fitting TAR model for the series. Future research will be devoted to better sorting out these conflicting results.

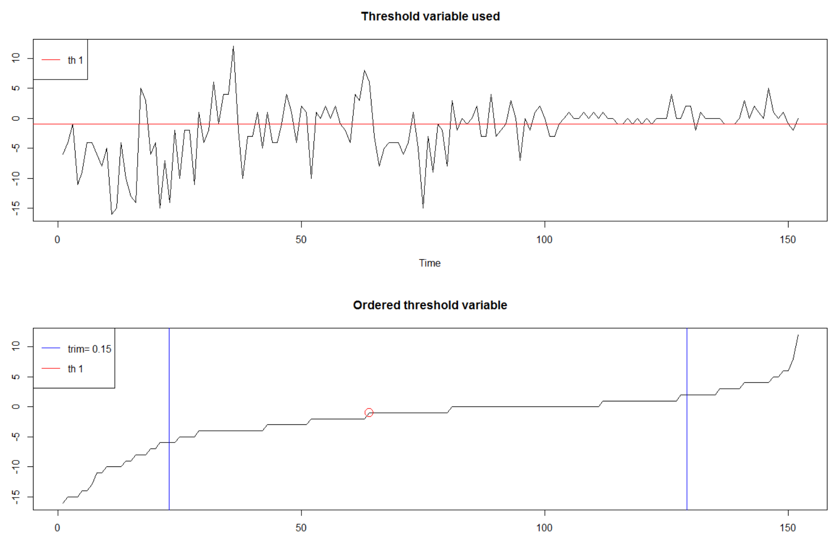

Figure 5: Regime Switching Plot for SETAR Model for Gallup based Measure of Macropartisanship

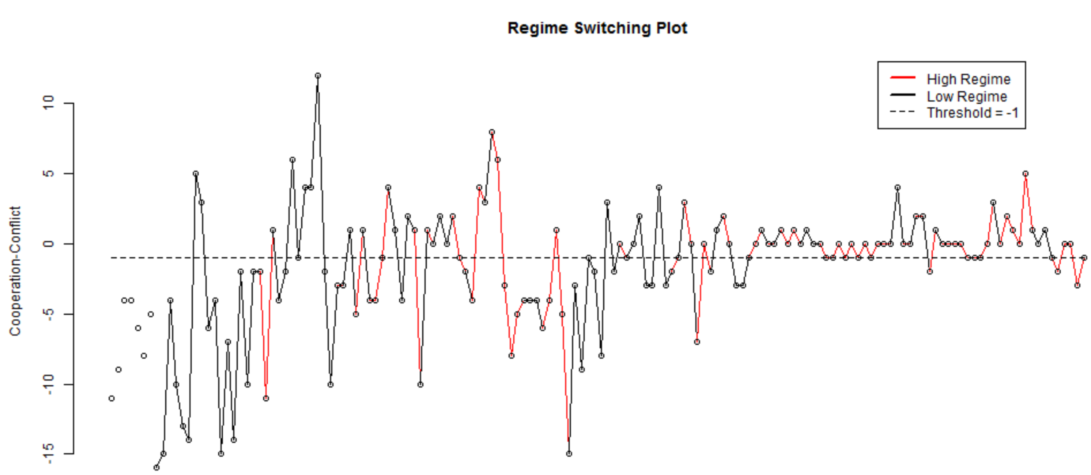

Figure 6: Regime Switching Plot for SETAR Model of Chinese Toward Taiwan Behavior

4 References

Balke, Nathan S. and Thomas B. Fomby. 1997. "Threshold Cointegration" *International Economic Review* 38(3):627-645.

Brandt, Patrick T., John R. Freeman, and Phil Schrodt. 2012. "Racing Horses: Constructing and Evaluating Forecasts in Political Sciences." Third draft presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

Brock, William, Davis Dechert, Jose Scheinkman, and Blake LeBaron. 1996. "A Test for Independence Based Upon the Correlation Dimension." *Econometric Reviews* 15: 197-235.

Chan, K.S., J.D. Petruccelli, H. Tong and S.W. Woolford. 1985 "A Multiple Threshold AR(1) Model." *Journal of Applied Probability* 22: 267-279.

Diks, Cees and Sebastiano Manzan. 2002. "Tests for Serial Independence and Linearity Based on Correlation Integrals." *Studies in Nonlinear Dynamics and Econometrics* 6(2).

Enders, Walter 2010. *Applied Econometric Time Series* Third Edition. John Wiley And Sons.

Enders, Walter, and C.W. J. Granger. 1998. "Unit-root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates." *Journal of Business and Economic Statistics* 16: 304-311.

Freeman, John R. 2012. "Research Note. Can Time Series Methods Be Used to Detect Path Dependence." Typescript. University of Minnesota. www.polisci.umn.edu/ freeman.

Freeman, John R. and John E. Jackson. 2012. "Symposium on Models of Path Dependence." *Political Analysis* 20(2):137-245.

Gerner, Deborah J., Philip A. Schrodt, and Ömür Yilmaz. 2009. "Conflict and Mediation Event Observations (CAMEO): An Event Data Framework for a Post Cold War World." In *International Conflict Mediation: New Approaches and Findings* J. Bercovitch and S. Gartner Editors. New York: Routledge.

Granger, C.W. J. and T. Teräsvirta. 1993 *Modelling Nonlinear Economic Relationships* NY: Oxford University Press.

Hansen, Bruce E. "Inference in TAR Models." 1997 *Studies in Nonlinear Dynamics and Econometrics* 2(1): 1-14.

Jackson, John and Kenneth Kollman. 2010. "A Formulation of Path Dependence With An Empirical Example." *Quarterly Journal of Political Science* 5(3): 257-289.

Jackson, John and Kenneth Kollman. 2012 "Modeling, Measuring, and Distinguishing Path Dependence, Outcome Dependence, and Outcome Independence." *Political Analysis* 20(2): 157-174.

Lasota, A. and M.C. Mackey. 1989. "Stochastic Perturbation of Dynamical Systems: The Weak Convergence of Measures." *Journal of Mathematical Analysis and Application* 138: 232-248.

Lin, Tse-min, Patrick Brandt, John Freeman and Phil Schrodt. In progress. "Understanding Cross-Strait Dynamics." Typescript. University of Minnesota.

Lütkepohl, Helmut. 1993. *Introduction to Multiple Time Series Analysis* Berlin: Springer-

Verlag.

Manzan, Sabastiano. 2003. "Essays in Nonlinear Economic Dynamics." Thela Thesis.

Pippenger, Michael K. and Gregory E. Goering. 1993. "A Note on the Empirical Power of Unit Root Tests under Threshold Processes." *Oxford Bulletin of Economics and Statistics* 55(4): 473-481.

Quinn, B.G. 1982. "Stationarity and Invertibility of Simple Bilinear Models." *Stochastic Processes and Their Applications* 12:225-230.

Tong, Howell. 1990. *Non-Linear Time Series: A Dynamical Systems Approach* NY: Oxford University Press.

5 Appendix

Exploratory Analysis of China Toward Taiwan Material Behavior Series, $C2TM_t$.

Figure 4 is the histogram of the data. This histogram suggests the series has a single mode.[18]

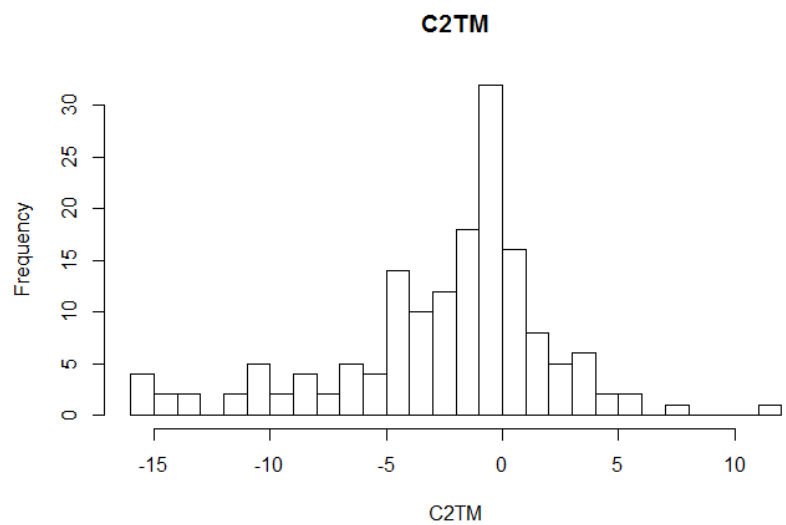

Figure 7: Histogram for C2TM(t).

Figures 5-8 are the nonparametric regression lines for what the R package tsDyn calls "Autopairs graphs." Curved lines–a hump, for example– suggests that a linear model may be inappropriate (see Tong 1990: 5.2.4, 7.2.3). With the exception of the plots for 12 and 16 lags, there seems to be little evidence of humps; the curves for 12 and 16 lags have only mild curvatures.[19]

[18]See Tong 1990: for a discussion for statistical tests for unimodality.

[19]The nonparametric regression lines are drawn with a function called "sm.regression" from the R library.

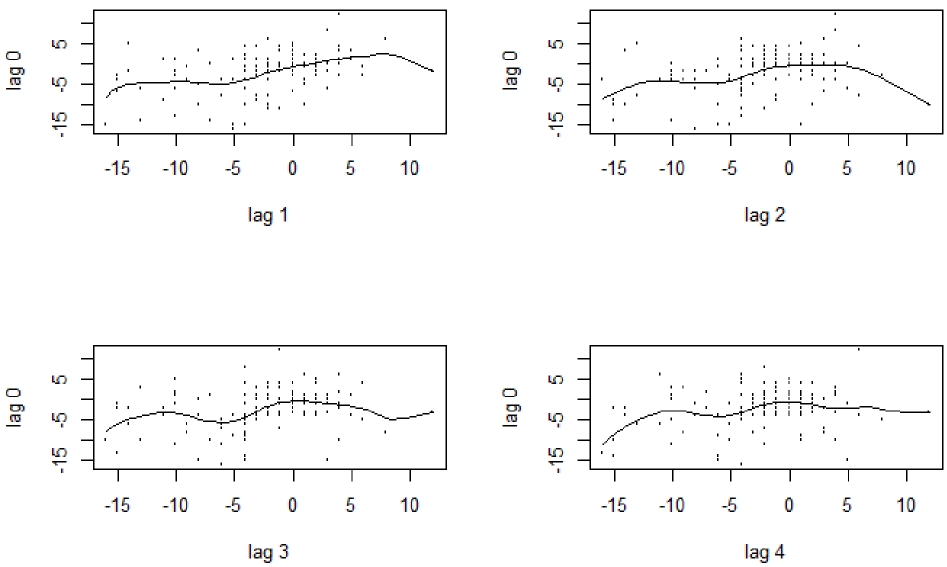

Figure 8: Autopairs Plot, Lag 1-4

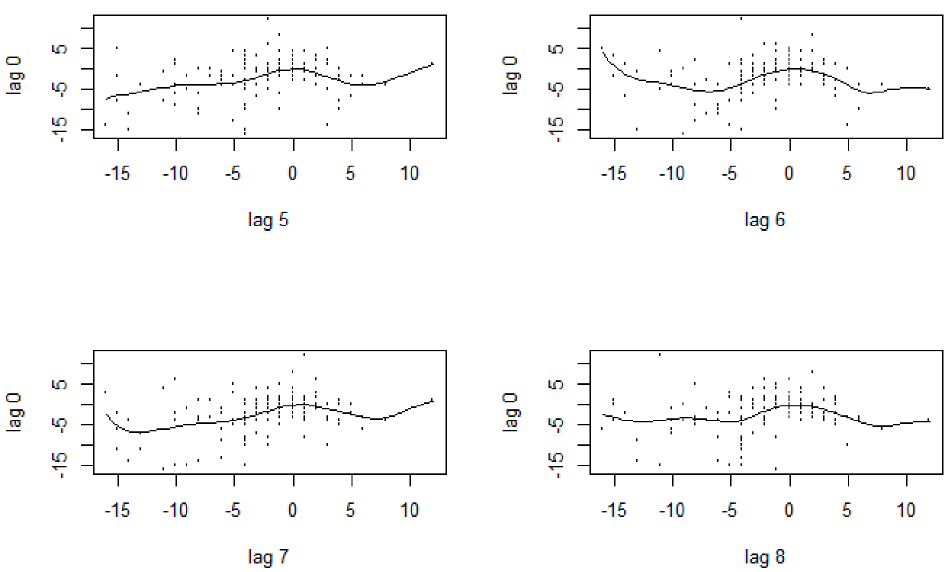

Figure 9: Autopairs Plot, Lag 5-8

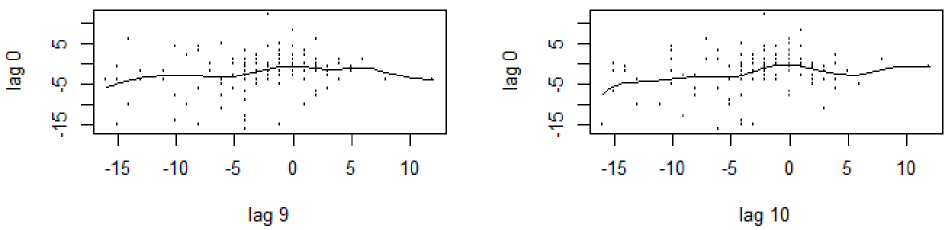

Figure 10: Autopairs Plot, Lag 9-12

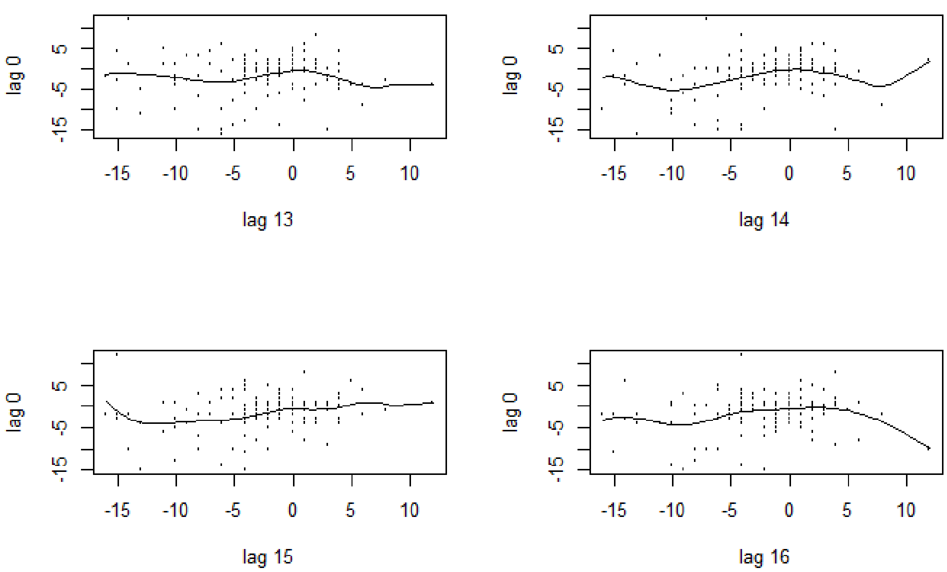

Figure 11: Autopairs Plot, Lag 13-16

5.1 Linear ARMA models for $C2TM_t$.

On the basis of an examination of the acf and pacf for the series, a variety of linear models were estimated. The results are reported in Tables 5 and 6. Briefly, for reasons that remain unclear, there is much evidence of a seventh order seasonality in $C2TM_t$. This characteristic of the series was best captured by an AR(1)-Seasonal AR(7) model with a constant.

5.2 The Issue of Stationarity

As noted in the text, the Dickey Fuller test assumes a linear, symmetric adjustment in a time series process. If the actual data generating process (DGP) is nonlinear, the Dickey Fuller test can produce mistaken inferences (Enders 2010: 477, Pippinger and Goering 1993). For this reason the results in Table 1 are not necessarily definitive. [Recall also that the KPSS test yielded ambiguous results (fn. 7)].

Enders (2010: Section 5.11) presents a test for unit roots under the alternative that the DGP is a simple TAR model. His analysis is an extension of work originally published in Enders and Granger (1993). Unfortunately, his test assumes a single variance for the time series of interest. Hence it is *not* applicable to the model we found with tsDyn. But, Enders' test is nonetheless informative. The TAR model used for Enders' test is:

$$\Delta C2TM_t = I_t\rho_1(C2TM_{t-1} - \tau) + (1 - I_t)\rho_2(C2TM_{t-1} - \tau) + \epsilon_t \tag{9}$$

where τ is the threshold and I_t is the indicator function here defined as:

$$I_t = \begin{cases} 1 & \text{if } C2TM_{t-1} \geq \tau \\ 0 & \text{if } C2TM_{t-1} < \tau. \end{cases}$$

Once more, this is not the model we found using tsDyn for $C2TM_t$ because it assumes, among other things, a single variance for Chinese net material behavior toward Taiwan. Now, if $\rho_1 = \rho_2 = 0$ the process would be a random walk; if we reject this restriction, we infer there is an "attractor" for $C2TM_t$ (it is stationary as long as $-2 < \rho_1, \rho_2 < 0$. The F statistic can be used to test for such an attractor but the critical values for this particular test are nonstandard. Enders supplies a Table (2010: 494) for critical values for the test of $\rho_1 = \rho_2 = 0$ for this TAR model. Should we reject the null hypothesis, we can proceed to test for asymmetric adjustment, $\rho_1 = \rho_2$. The standard F statistic and critical values can be used for this second restriction (these critical values are an approximation of the actual critical values which can be generated by Hansen bootstrap method).[20]

We implemented the test in Enders (2010) using the RATS code supplied with his book. The estimated model for the $C2TM_t$ series is:

[20]Enders and Granger(1993) analyze somewhat different TAR models. They also advocate a four step procedure which differs from the procedure outlined in Enders (2010: 479). In particular, in the latter one *starts* by finding the threshold of the TAR model rather than searching for this threshold *after* testing that $\rho_1 = \rho_2 = 0$. Table G in Enders (2010) does not appear in Enders and Granger (1993) apparently because the model used for the test in Enders (2010) is different from those used in the earlier article.

Coefficients and Statistics	AR(1)	AR(2)	AR(3)	AR(7)	ARMA(1,1)	MA(1)	MA(7)
AR(1)	0.42(.07)	.34(.08)	.32(.08)	.32(.08)	.95(.05)		
AR(2)		.19(.08)	.13(.08)	.07(.08)			
AR(3)			.15(.08)	.09(.08)			
AR(4)				.04(.08)			
AR(5)				.13(.08)			
AR(6)				-.15(.08)			
AR(7)				.28(.08)			
MA(1)					-.75(.10)	.32(.07)	.33(.08)
MA(2)							.18(.09)
MA(3)							.15(.08)
MA(4)							.14(.08)
MA(5)							.18(.08)
MA(6)							.02(.09)
MA(7)							.24(.09)
Constant	-2.03(.58)	-2.07(.70)	-2.11(.81)	-2.31(1.28)	-2.41(1.43)	-2.01(.46)	-2.08(.70)
σ^2	18.08	17.46	17.05	15.17	16.51	19.1	16.03
AIC	917.68	914.22	912.54	902.63	905.66	926.36	911.08
Q Stat(χ^2)	80.05	59.08	56.85	29.38	51.98	110.07	33.87
Signif	$p < .001$	p=.005	p=.006	p=.45	p=.03	$p < .001$	p=.24

Table 5: Estimations of Alternative Linear Models for $C2TM_t$. Numbers in parentheses are standard errors. Q's are Box-Ljung statistics for 36 lags adjusted for the number of the parameters in each model.

Coefficients and Statistics	MA(7) Restrictions	MA(1) Seasonal MA(6)	AR(7) Restrictions	AR(1) Seasonal AR(7)	AR(1) Seasonal AR(6)
AR(1)			.37(.07)	.38(.07)	.48(.08)
AR(7)			.30(.07)		
Seasonal AR				.32(.08)	-.14(.09)
MA(1)	.34(.08)	.33(.07)			
MA(2)	.16(.08)				
MA(7)	.24(.07)				
Seasonal MA		-.01(.07)			
Constant	-2.04(.56)	-2.00(.45)	-2.20(.93)	-2.10(.75)	-2.01(.57)
σ^2	16.96		16.11	16.21	17.79
AIC	911.97		902.10	903.05	917.30
Q Stat(χ^2)	49.36	111.05	39.95	34.66	66.48
Signif	p=.03	$p < .001$	p=.223	p=.436	p=.002

Table 6: Estimations of Alternative Linear Models for $C2TM_t$. Numbers in parentheses are standard errors. Q's are Box-Ljung statistics for 36 lags adjusted for the number of the parameters in each model.

$$\Delta C2TM_t = \underset{(1.53)}{1.10} - \underset{(2.73)}{.34} I_t(C2TM_{t-1}+6) - \underset{(-4.01)}{.79}(1-I_t)(C2TM_{t-1}+6) - \underset{(-2.55)}{1.80}\Delta C2TM_{t-1} + \epsilon_t$$

where the $\Delta C2TM_{t-1}$ term on the right side of the equation, as in Enders illustration, is included to account for any serial correlation in the errors of the equation and the numbers under the coefficients are t statistics.[21]. The estimated I_t, the indicator function is:

$$I_t = \begin{cases} 1 & \text{if } C2TM_{t-1} \geq -6 \\ 0 & \text{if } C2TM_{t-1} < -6 \end{cases}$$

The test for $\rho_1 = \rho_2 = 0$ yields a F(2, 153) of 18.07 which far exceeds the respective critical value in Enders' table G (approximately 6.21). This indicates that, according to the alternative of a simple TAR model with a single variance, the hypothesis of a unit root can be rejected.[22]

Thus, both the results of the Dickey Fuller test in Table 1 and this particular unit root test both produce the same conclusion: the $C2TM_t$ series is stationary. Future research will be devoted to finding a unit root test that has a TAR model with multiple variances as the alternative and to understanding the implications of nonstationarity for nonlinearity testing.

[21] The Durbin Watson statistic for the fitted TAR model is 2.03

[22] The F statistic for the equality of the coefficients in the TAR model here is F(1,153) = 2.92. This value has an exact statistical significance of .089 although it is only an approximation of the level of statistical significance.