A note on Structural Break, Nonlinear Time Series, Outliers, Nonlinear Granger Causality and extended Granger Causality for Nonlinear Time Series

Jiachuan Tian

Part I. Introduction

Structural Break, Nonlinearity and Outliers happen frequently in economic time series and they all in a sense appear to be Nonlinearity. By definition (Koop, 2002), a model where the dynamics change permanently in a way that cannot be predicted by the history of the series is considered as structural break. Alternatively, it is possible that dynamic properties can vary over the business cycle. And we refer to models which allow for dynamics which vary over the business cycle in a predictable way as "nonlinear" models. Still another possibility is that apparent departures from linearity are due to unpredictable large shocks which have only temporary effects, which is "outlier". Consider a three regime switching specification:

$$Y_t = \begin{Bmatrix} \alpha_{00} + \beta_{0p}(L)Y_{t-1} + \varepsilon_{0t} & if\ I_t = 0 \\ \alpha_{10} + \beta_{1p}(L)Y_{t-1} + \varepsilon_{1t} & if\ I_t = 1 \\ \alpha_{20} + \beta_{2p}(L)Y_{t-1} + \varepsilon_{2t} & if\ I_t = 2 \end{Bmatrix}$$

$$where\ \ I_t\ an\ indicator\ for\ the\ regimes\ and\ \beta_{ip}(L)\ a\ polynomial\ of\ order\ p\ in\ the\ lag\ operator.$$

We consider four ways of defining I_t

1. **A linear AR model** is obtained if $I_t = 0\ for\ \forall t.$
2. **A nonlinear model (TAR)** is obtained if
$$\begin{cases} I_t = 0 \quad if\ r_2 \le Y_{t-d} \le r_1 \\ I_t = 1 \quad if\ Y_{t-d} > r_1 \\ I_t = 2 \quad if\ Y_{t-d} < r_2 \end{cases}$$
3. **A structural Break model** is obtained if
$$\begin{cases} I_t = 0 \quad if\ t < \tau_1 \\ I_t = 1 \quad if \tau_1 \le t < \tau_2 \\ I_t = 2 \quad if\ t \ge \tau_2 \end{cases}$$
4. **A outlier model** is obtained if

$$\begin{cases} I_t \neq 0 & if\ t = \tau_1, \tau_2 \\ I_t = 0 & otherwise \end{cases}$$

$$and$$

$$\begin{cases} \beta_{0p}(L) = \beta_{1p}(L) = \beta_{2p}(L) \\ \varepsilon_{0t} = \varepsilon_{1t} = \varepsilon_{2t} \\ \alpha_{00} \neq \alpha_{10} \neq \alpha_{20} \end{cases}$$

In specifying model 1-4, Koop (2000) suggest a Bayesian method, while Giordani, Kohn and Dijk suggest a State-space framework, among many others. Now we will give a closer look on each model and methods above.

Part II. Structural Break

1. Motivation:

"Parameter instability for economic models is a common phenomenon. And this is particularly true for time series data covering an extended period, as it is more likely for the underlying data-generating mechanism to be disturbed over a longer horizon by various factors such as policy-regime shift."

2. Core Topics in structural breaks:

a. Estimation and inference about break dates for single equations and multiple equations;

b. Tests for a single structural break and multiple structural breaks ;

c. Tests for unit root in the presence of structural breaks;

d. Tests for cointegration in the presence of structural breaks.

3. Estimation and inference about break dates

Bai gave the estimation and inference about one single break point(1995). Perron and Bai (1998) later introduced the estimation and inference about multiple break points. Here we only talk about the multiple break points situation.

3.a Model and Assumptions:

For a linear regression,

$$y_t = x_t'\beta + z_t'\delta_j + \epsilon_t \quad t = T_{j-1} + 1 \dots . T_j$$

for *j=1...(m+1)* denoting there are *m* structural breaks (or *m+1* regimes). y_t denotes the observation at time t, x_t and z_t are vectors of regressors. And such regression can be rewritten in matrix form as

$$Y = X\beta + Z_0\delta + \epsilon$$

For each *m* partition (*T1...Tm*), the associated least-squares estimates of β and δj are obtained by minimizing the sum of squared residuals $\epsilon'\epsilon$. Let $\hat{\beta}(\{Tj\})$ and $\hat{\delta}(\{Tj\})$ denote the estimates based on the given *m* partition *(T1, ..., Tm)* denoted {Tj}. Substituting these in the objective function and denoting the resulting sum of squared residuals as *ST (T1, ..., Tm)*, the estimated break points are

$$\left(\widehat{T_1}, \dots \widehat{T_m}\right) = argmin_{(T_1,\dots T_m)} S_T(T_1 \dots T_m)$$

The assumptions of such model relax from iid models up to a shift [Yao, 1987], [Bhattacharya, 1987] to a mean shift for a Gaussian AR proceess [Picard, 1994], and further to multiple regressions [Bai,1995]. The model has assumptions on regressors, errors, break dates, and the minimization procedure. The specific assumptions refer to *Change Point Estimation in multiple regression models* [Bai,1995] and *Dealing with structural break* [Perron, 2005].

3.b Consistency and Asymptotic Distribution of the break date estimators $\left(\widehat{T_1}, \dots \widehat{T_m}\right)$

"With the assumptions on the regressors, the errors and given the asymptotic framework adopted, the limit distributions of the estimates of the break dates are independent of each other. Hence, for each break date, the analysis becomes exactly the same as if a single break has occurred." This reduces the study of asymptotic distribution in multiple break points to single one, which has well been stated by Bai (1995).
Under the assumptions,

$$\widehat{T_1} = T_1 + O_p(\|\delta_T\|^{-2})$$

And with the results, together with $\epsilon_t$being uncorrelated and $E\epsilon_t^2$=σ^2 for all *t*, then

$$\begin{bmatrix} \sqrt{T}(\hat{\beta} - \beta) \\ \sqrt{T}(\hat{\delta} - \delta_T) \end{bmatrix} \xrightarrow{d} N(0, \sigma^2 V^{-1})$$

where $V = plim \frac{1}{T} \begin{bmatrix} \sum_{t=1}^{T} x_t x_t' & \sum_{t=T1}^{T} x_t z_t' \\ \sum_{t=T1}^{T} z_t x_t' & \sum_{t=T1}^{T} z_t z_t' \end{bmatrix}$.

And the under assumptions,

$$\hat{T} - T \xrightarrow{d} arg \max_m W^*(m)$$

where $W^*(m) = \begin{cases} 0, m = 0 \\ W_1(m), m < 0 \\ W_2(m), m > 0 \end{cases}$, a two sided random walk with (stochastic) drift.

4. Tests for a single structural break and multiple structural breaks

4.a CUSUM Test [Brown, Durbin and Evans, 1975]
For a linear regression with k regressors

$$y_t = x_t'\beta + \epsilon_t$$

The CUSUM statistic is defined as

$$CUSUM = \max_{k+1<r\leq T} \left|\frac{\sum_{t=k+1}^{r} \widetilde{v_t}}{\hat{\sigma}\sqrt{T-k}}\right| /(1 + 2\frac{r-k}{T-k})$$

Where $\hat{\sigma}^2$ is a consistent estimate of the variance of ϵ_t.

$$\widetilde{v_t} = \frac{y_t - x_t'\widehat{\beta_{t-1}}}{f_t} \text{ and } f_t = (1 + x_t'(X_{t-1}'X_{t-1})x_t)^{1/2}$$

The asymptotic distribution of CUSUM is:

$$CUSUM \xrightarrow{d} sup_{0\leq r\leq 1}\left|\frac{W(r)}{1+2r}\right|$$

Where *W(r)* is a unit Wiener process defined on (0,1). [Sen, 1982]

4.b CUSSQ Test [Brown, Durbin and Evans, 1975]

$$CUSSQ = \max_{k+1<r\leq T}\left|S_T^{(r)} - \frac{r-k}{T-k}\right|$$

Where $S_T^{(r)} = \left(\sum_{t=k+1}^{r} \widetilde{v_t}^2\right)/(\sum_{t=k+1}^{T} \widetilde{v_t}^2)$

Ploberger and Kramer (1990) considered the local power functions of the CUSUM and
CUSUM of squares. The former has non-trivial local asymptotic power unless the mean
regressor is orthogonal to all structural changes. On the other hand, the latter has only trivial local power (i.e., power equal to size) for local changes that specify a one-time change in the coefficients (see also Deshayes and Picard, 1986). This suggests that the CUSUM test should be preferred, a conclusion we shall revisit below.

4.c *sup-LR* test

The *sup-LR* test statistic is:

$$sup_{\lambda_1\in\Lambda_\epsilon} LR_T(\lambda_1)$$

Where $LR_T(\lambda_1)$ denotes the value of likelihood ratio evaluated at some break point $T_1 = [T\lambda_1]$ and maximization is restricted over break fractions that are in $\Lambda_\epsilon = [\epsilon_1, 1-\epsilon_2]$.

The asymptotic distribution of *sup-LR* test is:

$$sup_{\lambda_1\in\Lambda_\epsilon} LR_T(\lambda_1) \xrightarrow{d} sup_{\lambda_1\in\Lambda_\epsilon} G_q(\lambda_1)$$

Where $G_q(\lambda_1) = \frac{[\lambda_1 W_q(1) - W_q(\lambda_1)]'[\lambda_1 W_q(1) - W_q(\lambda_1)]}{\lambda_1(1-\lambda_1)}$

$W_q(\lambda)$ a vector of independent Wiener processes of dimension q, the number of coefficients that are allowed to change.

4.d *sup-Wald* test

The *sup-Wald* test statistic is defined as

$$sup_{\lambda_1 \in \Lambda_\epsilon} W_T(\lambda_1; q)$$

Where$W_T(\lambda_1) =$
$[SSR(1,T) - SSR(1,T_1) - SSR(T_1+1,T)]/\{[SSR(1,T_1) + SSR(T_1+1,T)]/T\}$

Where *SSR(i,j)* is the sum of squared residuals from regressing y_t on a constant using data from data *i* to date *j, i.e.*

$$SSR(i,j) = \sum_{t=i}^{j}\left(y_t - \frac{1}{j-i}\sum_{t=i}^{j} y_t\right) = \sum_{t=i}^{j}(e_t - \sum_{t=i}^{j} e_t)$$

The asymptotic distribution of *sup-Wald* test statistic is:

$$W_T(\lambda_1) \xrightarrow{d} \frac{1}{\lambda_1(1-\lambda_1)}[\lambda_1 W(1) - \lambda_1 W(\lambda_1) - (1-\lambda_1)W(\lambda_1)]^2$$

Note here $W_T(k)$ is monotonic transformation of $S_T(k)$. So it follows that

$$\left(\widehat{T_1}, \dots \widehat{T_m}\right) = argmin_{(T_1,\dots T_m)} S_T(T_1 \dots T_m) = argmin_{(T_1,\dots T_m)} W_T(T_1 \dots T_m)$$

Hence, the estimator obtained by minimizing the sum of squared residuals is the same as maximizing Wald-type statistics.

5. Tests for unit root in the presence of structural breaks

Consider a linear regression model in which we allow shifts in both intercept and slope,

$$y_t = \mu_1 + \beta_1 t + (\mu_2 - \mu_1)DU_t + (\beta_2 - \beta_1)DT_t^* + \epsilon_t$$

where $DU_t = 1, DT_t^* = t - T_1 \; if \; t > T_1 \; and \; 0 \; otherwise$

(Such model is call model AO-C by Perron)

Rewriting the model AO-C

$$y_t = \mu + \theta DU_t + \beta t + \gamma DT_t^* + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i} + e_t$$

Then the test statistic for a unit root allowing for changes at unknown dates is:

$$t_\alpha^* = inf_{\lambda_1 \in [\epsilon, 1-\epsilon]} t_\alpha(\lambda_1)$$

where $t_\alpha(\lambda_1)$ is the t-statistic for testing $\alpha = 1$

The asymptotic distribution of such statistic is:

$$t_\alpha^* \xrightarrow{d} inf_{\lambda_1 \in [\epsilon, 1-\epsilon]} \frac{\int_0^1 W^*(r,\lambda_1) dW(r)}{[\int_0^1 W^*(r,\lambda_1)^2 dr]^{1/2}}$$

$W^*(r,\lambda_1)$ is the residual function from a projection of a Wiener process *W(r)* on the relevant continuous time versions of the deterministic components.

Part III. Nonlinear Time Series Models

1. **Test for nonlinearity**

The BDS test developed by Brock, Dechert and Scheinkman (1987) is arguably the most popular test for nonlinearity. When applied to the residuals of a fitted linear time series model, the BDS test can be used to detect remaining dependence and the presence of omitted nonlinear structure (Zivot, 2005). The null hypothesis of BDS test is that the linear model specification is appropriate (residual is i.i. distributed).

Consider a time series $x_t, t = 1,2, \dots, T$ and define its *m*-lag history as $x_t^m = (x_t, x_{t-1}, \dots x_{t-m+1})$. The correlation integral at embedding dimension m can be estimated by:

$$C_{m,\varepsilon} = \frac{2}{(T-m+1)(T-m)} \sum_{m \le s < t \le T} \sum I(x_t^m, x_s^m; \varepsilon)$$

$$where\ I(x_t^m, x_s^m; \varepsilon) = \begin{cases} |x_{t-i} - x_{s-i}| < \varepsilon & for\ i = 0,1, \dots, m-1 \\ 0 & otherwise \end{cases}$$

The correlation integral estimates the probability that any two m-dimensional points are within a distance of $\varepsilon\ of\ each\ other,$ i.e.

$$\widetilde{\Pr}(|x_t - x_s| < \varepsilon, |x_{t-1} - x_{s-1}| < \varepsilon, \dots, |x_{t-m+1} - x_{s-m+1}| < \varepsilon) = C_{m,\varepsilon}$$

If $x_t \sim iid,$

$$C_{1,\varepsilon}^m = \Pr(|x_t - x_s| < \varepsilon)^m$$

Then the BDS statistic is defined as follows:

$$V_{m,\varepsilon} = \sqrt{T} \frac{C_{m,\varepsilon} - C_{1,\varepsilon}^m}{S_{m,\varepsilon}}$$

$$where\ S_{m,\varepsilon} = SD(\sqrt{T}(C_{m,\varepsilon} - C_{1,\varepsilon}^m))$$

$$Then\ V_{m,\varepsilon} \xrightarrow{d} N(0,1)$$

2. **Markov Switching Model (Hamilton, 1989)**

Consider a general VAR time series model:

$$z_t = \mu + D_t + S_t + \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t$$

$$where$$

$$D_t\ is\ the\ trend\ conponent\ and\ S_t\ the\ seasonal\ component$$

D_t obeys a Markov trend in levels if

$$D_t = \alpha_1 . x_t + D_{t-1}\ where$$

$$x_t = 0\ or\ 1\ denotes\ the\ unobserved\ state\ of\ the\ system$$

We assume the transition between states is governed by a first-order Markov process:

$$\Pr(X_t = 1 | X_t = 1) = p,$$
$$\Pr(X_t = 0 | X_t = 1) = 1 - p,$$
$$\Pr(X_t = 0 | X_t = 0) = q,$$
$$\Pr(X_t = 1 | X_t = 0) = 1 - q.$$

Then the model can be called a VAR with Markov trend in levels. Further., Hamilton (1989) develop a time series model with Markov trend in lags and develop estimation algorithm and filter for such model.

3. **TAR and SETAR**

A Threshold Autoregressive model (TAR) (Potter,1995) is defined as that in Part I:

$$Y_t = \begin{Bmatrix} \alpha_{00} + \beta_{0p}(L)Y_{t-1} + \varepsilon_{0t} & if\ I_t = 0 \\ \alpha_{10} + \beta_{1p}(L)Y_{t-1} + \varepsilon_{1t} & if\ I_t = 1 \\ \alpha_{20} + \beta_{2p}(L)Y_{t-1} + \varepsilon_{2t} & if\ I_t = 2 \end{Bmatrix}$$

$$where\ \ I_t\ an\ indicator\ for\ the\ regimes\ and\ \beta_{ip}(L)\ a\ polynomial\ of\ order\ p\ in\ the\ lag\ operator. And\ I_t\ is\ defined\ as$$

$$\begin{cases} I_t = 0 \quad if\ r_2 \le z_t \le r_1 \\ I_t = 1 \quad if\ z_t > r_1 \\ I_t = 2 \quad if\ z_t < r_2 \end{cases}$$

$$where\ z_t\ is\ a\ weakly\ exogenous\ threshold\ vairable$$

So in each regime, the time series Y_t follows a different AR(*p)* model.
When the threshold variable $z_t =$
$Y_{t-d}\ with\ the\ delay\ parameter\ d\ being\ a\ positive\ integrer,$
then the dynamics or regime of Y_t is determined by its own lagged value Y_{t-d} is called a self-exciting **TAR (SETAR)** model.

4. **STAR**

In the TAR models, a regime switch happens when the threshold variable crosses a certain threshold. However, it suggests the regime switch is discontinuous. If the discontinuity of the thresholds is replaced by a smooth transition function, TAR models can be generalized to **smooth transition autoregressive (STAR)** model.

Take exponential STAR **(ESTAR)** model as an example. Replace z_t by $G(z_t)$, a smooth transition function 0< $G(z_t)$<1 which depends on a transtition variable z_t (like the threshold variable in TAR model), the model becomes a smooth transition model:

$$Y_t = \mu + \beta_{0p}(L)Y_{t-1}\big(1 - G(z_t)\big) + \beta_{1p}(L)Y_{t-1}G(z_t) + \varepsilon_t$$

$$where\ G(z_t; \gamma, c) = \ 1 - e^{-\gamma(z_t - c)^2}, \gamma > 0\ (exponential)$$

To avoid issues caused by the unidentified STAR model parameters under the null hypothesis of a linear AR model, Luukkonen, Saikkonen and Terasvirta (1988) propose to replace the transition function G(z_t; γ, c) by a suitable Taylor series approximation around γ = 0.Then the test for ESTAR nonlinear can be conducted within the linear context.

Part IV. Nonlinear Granger Causality

1. (Linear) Granger causality test

Granger causality test (Granger, 1969) is designed to detect causal direction in a bivariate linear time series. It tests on the correlation between the current value of one variable and the past values of the other one:

Consider a bivariate VAR(p):

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} a_{10} \\ \vdots \\ a_{(k+l)0} \end{bmatrix} + \sum_{h=1}^{p} \begin{bmatrix} a_{11,t-h} & \cdots & a_{1(k+l),t-h} \\ \vdots & \ddots & \vdots \\ a_{(k+l)1,t-h} & \cdots & a_{(k+l)(k+l),t-h} \end{bmatrix} \begin{bmatrix} x_{t-h}^1 \\ x_{t-h}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{(k+l),t} \end{bmatrix}$$

$$where\ \Sigma = \begin{bmatrix} \sigma_{1,t}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{(k+l),t}^2 \end{bmatrix}.$$

To test Ho:$[x_t^2]_j$does not Granger-cause $[x_t^1]_i$, we need to jointly test on $a_{ij,h} = 0\ for\ h = 1,2,\dots,p$. And the alternative is: $\exists a_{21,i}\ ne\ 0\ for\ some\ i.$ The test statistic is distributed as $F(p,(k+l),(k+l) \times p - A)$, where *A* equals to the total number of parameters in the above VAR (p) including the deterministic regressors.

2. Nonlinear Granger causality test

The linear causality tests, such as the Granger test (1969), can fail to uncover nonlinear predictive power (Baek and Brock, 1992; Hiemstra and Jones, 1994). Hiemstra and Jones (1994), based on Baek and Brock (1992), proposed a nonparametric statistical method based on the correlation integral to detect nonlinear causal relations between time series. Consider two strictly stationary and weakly dependent time series$\begin{bmatrix} x_t \\ y_t \end{bmatrix}$:

$$x_t^m = (x_t, x_{t+1}, \dots, x_{t+m-1}), m = 1,2,\dots, t = 1,2,\dots$$
$$x_{t-lx}^{lx} = (x_{t-lx}, x_{t-lx+1}, \dots, x_{t-1}), lx = 1,2,\dots, t = lx+1, lx+2,\dots$$
$$y_{t-ly}^{ly} = \left(y_{t-ly}, y_{t-ly+1}, \dots, y_{t-1}\right), ly = 1,2,\dots, t = ly+1, ly+2,\dots$$

The then null hypothesis is that Y does not strictly Granger cause X, which by definition means:

$$\Pr\left(\|x_t^m - x_s^m\| < e \middle| \left\|x_{t-lx}^{lx} - x_{s-lx}^{lx}\right\| < e, \left\|y_{t-ly}^{ly} - y_{s-ly}^{ly}\right\| < e\right)$$

$$= \Pr\left(\|x_t^m - x_s^m\| < e \middle| \left\|x_{t-lx}^{lx} - x_{s-lx}^{lx}\right\| < e\right)$$

Rewrite this condition,

$$\frac{C_1(m+lx, ly, e)}{C_2(lx, ly, e)} = \frac{C_3(m+lx, e)}{C_4(lx, e)}$$

$$where$$

$$C_1(m+lx, ly, e) = \Pr(\left\|x_{t-lx}^{m+lx} - x_{s-lx}^{m+lx}\right\| < e, \left\|y_{t-ly}^{ly} - y_{s-ly}^{ly}\right\| < e)$$

$$C_2(lx, ly, e) = \Pr(\left\|x_{t-lx}^{lx} - x_{s-lx}^{lx}\right\| < e, \left\|y_{t-ly}^{ly} - y_{s-ly}^{ly}\right\| < e)$$

$$C_3(m+lx, e) = \Pr(\left\|x_{t-lx}^{m+lx} - x_{s-lx}^{m+lx}\right\| < e)$$

$$C_4(lx, e) = \Pr\left(\left\|x_{t-lx}^{lx} - x_{s-lx}^{lx}\right\| < e\right)$$

As in BDS test, we use correlation-integral to estimate these probabilities:

$$C_1(m+lx, ly, e, n) = \frac{2}{n(n-1)}\sum\sum I(x_{t-lx}^{m+lx}, x_{s-lx}^{m+lx}, e) \cdot I(y_{t-ly}^{ly}, y_{s-ly}^{ly}, e)$$

$$C_2(lx, ly, e, n) = \frac{2}{n(n-1)}\sum\sum I(x_{t-lx}^{lx}, x_{s-lx}^{lx}, e) \cdot I(y_{t-ly}^{ly}, y_{s-ly}^{ly}, e)$$

$$C_3(m+lx, e) = \frac{2}{n(n-1)}\sum\sum I(x_{t-lx}^{m+lx}, x_{s-lx}^{m+lx}, e)$$

$$C_4(lx, e) = \frac{2}{n(n-1)}\sum\sum I(x_{t-lx}^{lx}, x_{s-lx}^{lx}, e)$$

$where\ t, s = \max(lx, ly) + 1, \ldots, T - m + 1;$

$n = T - \max(lx, ly) - m + 1$

For given values$m, lx, ly \geq 1\ and\ e > 0$, under the assumptions that $x_t\ and\ y_t$ are strictly stationary, weakly dependent and ergodic, if y_t does not strictly Granger cause x_t, then

$$\sqrt{n}\left[\frac{C_1(m+lx, ly, e)}{C_2(lx, ly, e)} - \frac{C_3(m+lx, e)}{C_4(lx, e)}\right] \sim N\left(0, \sigma^2(m, lx, ly, e)\right)$$

The statistics here will be applied to the residuals from the VAR models. And the null hypothesis will be rejected at 5% if the statistic is bigger than 1.96, as that in BDS test.