Time Series Notes

By JiachuanTian

1.Preliminary

(i) White Noise

A continuous time stochastic process $w(t) where\ t \in \mathbb{R}$ is a white noise process if and only if its mean function and autocorrelation function satisfy the following[1]:

$$E\{w(t)\} = 0$$

$$E\{w(t_i)w(t_j)\} = 0\ for\ \forall i \neq j$$

$$E\{w(t_i)w(t_i)\} = \sigma^2 < \infty$$

(ii) Cross-sectional data, longitudinal data and panel data[2]

>Cross-sectional data refers to data collected by observing many subjects (such as individuals, firms or countries/regions) at the same point of time, or without regard to differences in time.

>Longitudinal data follows one subject's changes over the course of time. Time series data is longitudinal data.

>Panel data (or time-series cross-sectional (TSCS) data), combines both cross sectional data and longitudinal data and looks at multiple subjects and how they change over the course of time. Panel analysis uses panel data to examine changes in variables over time and differences in variables between subjects.

2.Data Generating Mechanism

>Define$\{X_t\}, t = t_0, t_1, \dots, t_\infty$ as a time series generated by a DGM, represented by $J(X_t)$ and$\{x_t\},\ t = t_0, t_1, \dots, t_T\}$ as a realization of $\{X_t\}$. In general, a time series sample $\{x_t\}$ is a sequence of finite dimension that is drawn from the joint distribution $J(X_t)$ and is a sample with sample size *T*.

The empirical issue of interest in time series modeling is to represent the dynamic evolution of $\{X_t\}$ based on a single, observed realization. This is fundamentally more challenging that the problem encountered in cross-sectional data modeling. to begin, because we have only a finite sample we must ask whether $J(x_t)_T$ is representative of $J(X_t)$ itself. This motivates the examination of stationarity.

3. Stationarity

(i) Strictly stationarity: a strictly stationary process is a stochastic process that can be represented as any finite joint distribution for any finite period T, such that joint probability distribution does not change when shifted in time or space. As for time series, if a time series $\{X_t\}, t = t_0, t_1, \dots, t_\infty$ is strictly stationary, $\{X_1 \dots X_T\}$ and $\{X_{1+n} \dots X_{T+n}\}$ a have the same joint distribution for $\forall$ *T* and *n*>0.

---

[1] Here the ti and tj notations only represent different observation time rather than a subsequence of time for such model is a Continuous Time stochastic process with uncountable time points.

[2] Refer to http://en.wikipedia.org/wiki/Cross-sectional_data.

If $J(X_t)$ is strictly stationary, then there exists $J_T(X_t|\theta)$ such that $j(x_t)$ is independent of time. Thus we can use $\{x_t\}$ to characterize $J_T(X_t|\theta)$.

(ii) And if a process is stationary, parameters such as mean and covariance function, if they exist, should be independent of time, by which the weak stationarity is induced.

>weakly stationary: Xt is weakly stationary if

a. $\mu(\mathrm{x_t})$ is independent of t

b. $\gamma_{\mathrm{x}}(t+h,t)$ is independent of t for each h

(iii) In the presence of non-stationary variable, there might be what Granger and Newbold (1974) call a spurious regression. A spurious regression appears to have significant relationship among variable but the results are in fact without any economic meaning. An example of spurious regression (Jesús Gonzalo) is the regression of US Defense Expenditure (Y) (1971-1990, annual data) on Population of South African (X): $\hat{Y}$= -368.99 + .0179 *X*, R^2= .940, Corr = .9694. The regression has a high R-square but no economic meaning.

4. Unit Root and Integration

(i)**Definition**

A linear stochastic process has a unit root if 1 is a root of the process's characteristic equation. Consider a discrete time stochastic process AR (p) (without drift and trend):

$$x_t = \sum_{1}^{p} a_i \Delta x_{t-i} + \varepsilon_t$$

If m = 1 is a root of the characteristic equation:

$$m^p - m^{p-1}a_1 - m^{p-2}a_2 - m^{p-3}a_3 \ldots . - a_p = 0,$$

where p is the lag length.

Then the stochastic process has a unit root or, alternatively, is integrated of order one, denoted I (1). If m = 1 is a root of multiplicity r, then the stochastic process is integrated of order r, denoted I(r).

(ii)Unit root and stationarity

If a process has a unit root, it is non-stationary. An I (1) stochastic process will be stationary after differenced once, and an I(r) stochastic process will stationary after differenced r times. In this sense unit root and stationarity are connected.

(iii)Integration

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series

data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. For example, an ARIMA(p,1,q) is integrated by order 1, or I(1); and the first difference of such process would be stationary.

5. Dickey-Fuller test

Assume a time series (with no trend and drift) can be modeled as:

$$x_t = a_1 x_{t-1} + \varepsilon_t, t = 1, \dots, T$$

Rewrite the equation we have,

$$\Delta x_t = (a_1 - 1)x_{t-1} + \varepsilon_t$$
$$\Delta x_t = \gamma x_{t-1} + \varepsilon_t$$

whereΔ is the first difference operator. Testing the hypothesis that there is unit root is to test $\gamma = 0$. Dickey and Fuller (1979) consider three different regression equations: with no drift and trend, with drift and with drift and trend. In each case the null hypothesis is Ho: $\gamma = 0\ (unitroot)$. The test statistics and critical values are summarized in the following table.

TABLE-1 ADF Test Model and Critical Values

Model	Hypothesis	Test Statistics	Critical Values for 95% and 99% C.I.
$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_2 t + \varepsilon_t$	$\gamma = 0$	τ_τ	-3.45 and -4.04
	$\gamma = \alpha_2 = 0$	ϕ_3	6.49 and 8.73
	$\alpha_0 = \gamma = \alpha_2 = 0$	ϕ_2	4.88 and 6.50

$$where \quad \phi_i = \frac{[SSR(restricted) - SSR(unrestricted)]/r}{SSR(unrestricted)/(T-k)} \sim F$$

$$\tau_\tau = \tau = \tau_\mu = \frac{\gamma - 0}{SE(\gamma)} \sim t$$

6. Augmented Dick-Fuller Test

If the error term has autocorrelation more than one period, the unit root test (with drift and trend) can be modified as

$$\Delta x_t = \alpha_0 + \gamma x_{t-1} + \sum_{2}^{k} \lambda_i \Delta x_{t-i+1} + \varepsilon_t$$

The unit root test is then carried out under Ho: $\gamma = 0\ (unitroot)$against the alternative hypothesis of Ha:$\gamma \neq 0\ (no\ unit\ root)$.And the test statistic is the same as that in DF test:

$$\tau = \frac{\gamma - 0}{SE(\gamma)} \sim t$$

The number of lagged difference terms (k) to be included can be chosen based on AIC[3] (Greene, 1993).

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price transmission by time series\codes\R code piece (CPI 25 goods)\stationarity test.

7. Phillips-Perron test

The Phillips-Perron test is also a unit root test. It builds on the Dickey-Fuller test of the null hypothesis $\gamma = 0$ in the model:

$$\Delta x_t = \gamma x_{t-1} + \varepsilon_t, t = 1, \dots, T$$

The test makes a non-parametric correction to the t-test statistic generated by the DF test. Let the t-statistic of the DF test is the same as mentioned above:

$$\tau = \frac{\gamma - 0}{SE(\gamma)} \sim T$$

Phillips and Perron (1988) considered a different way to handle short-run memory suggested by moving average components and advocated non-parametric corrections of the test statistics:

$$Z(\hat{\gamma}) = T(\hat{\gamma}) - \frac{\widehat{\Omega} - \widehat{\Gamma}(0)}{2T^{-2}\sum x_{t-1}^2},$$

$$Z(\tau) = \frac{\sqrt{T^{-1}\sum \widehat{\varepsilon_t}^2}}{\sqrt{\widehat{\Omega}}}\tau - \frac{\widehat{\Omega} - \widehat{\Gamma}(0)}{2\sqrt{\widehat{\Omega}}\sqrt{T^{-2}\sum x_{t-1}^2}}$$

where$\widehat{\Gamma}(0) and \widehat{\Omega}$ are consistent estimators of the variance and long-run variance of ε_t respectively.

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price tramsmission by time series\codes\R code piece (CPI 25 goods)\stationaritytest.r.

8. Co-integration: Definition

If two or more series are individually integrated (in the time series sense) but some linear combination of them has a lower order of integration, then the series are said to be co integrated. If two series are both I(1), there might be a linear combination of levels of integrated variables such that the error in the relationship is stationary. Such variables are

[3]In the general case, the AIC is: $AIC = 2K - 2\ln(L)$ where K is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model.

said to be co-integrated. That is, they share a common unit root and the sequence of stochastic shocks is common to both. Co-integration is a powerful concept that allows capturing of any equilibrium relationship even between non-stationary series within a stationary model. Co-integration implies that prices move closely together in the long-run, although in the short-run they may drift apart.

To test co-integration, Granger and Engle (1987) developed a simple procedure which comprises of estimating the static co-integration regression, and apply unit test, such as the ADF and PP to the estimated residuals, in order to test the null hypothesis of no co-integration. Let $x_t = \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}_{(k+l)\times 1}$

$$[x_t^1]_i = b_0 + b_1[x_t^2]_j + \nu_t$$

If $[x_t^1]_i, i = 1, \dots, k$ and $[x_t^2]_j, j = 1, \dots, l$ are both integrated of order one, and ν_t is I(0), then $[x_t^1]_i$ and $[x_t^2]_j$ are co-integrated. The test procedure follows the Dickey-Fuller test.

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price transmission by time series\codes\R code piece (CPI 25 goods)\cointegration.r

9. Vector Autoregression

Vector auto-regression (VAR) is an econometric model used to capture the evolution and the interdependencies between multiple time series, generalizing the univariate AR models. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model.[4] A structural VAR(p) could take the form:

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

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

Note that all the vectors included in the VAR structure should be stationary.

The traditional econometric models structure on the parameters $a_{ij\,1,t}$ often prior to estimation. In contrast, use of a general, unrestricted VAR provides a starting point from which structural restrictions can be examined as hypotheses.

[4] Refer to http://en.wikipedia.org/wiki/Vector_autoregression.

10. Vector error correction model (VECM)

As for the motivation of introducing Augmented Dickey-Fuller test, if the error term has autocorrelation of more than one period, the VAR can be modified as Vector error correction model (VECM). A vector error correction model is a dynamic model in which the movement of the variables in any period is related to the previous period's gap from long-run equilibrium. AVECM (p) form is written as the first difference of the VAR above:

$$\Delta x_t = \begin{bmatrix} \Delta\, x_t^1 \\ \Delta\, x_t^2 \end{bmatrix} =$$

$$= \begin{bmatrix} \gamma_{10,t} \\ \vdots \\ \gamma_{(k+l)0,t} \end{bmatrix}^T \begin{bmatrix} x_{t-1}^1 \\ x_{t-1}^2 \end{bmatrix} + \sum_{i=2}^{p} \begin{bmatrix} \lambda_{11,t-i} & \cdots & \lambda_{1(k+l),t-i} \\ \vdots & \ddots & \vdots \\ \lambda_{(k+l)1,t-i} & \cdots & \lambda_{(k+l)(k+l),t-i} \end{bmatrix} \begin{bmatrix} \Delta x_{t-i}^1 \\ \Delta x_{t-i}^2 \end{bmatrix} + \begin{bmatrix} \nu_{1,t} \\ \vdots \\ \nu_{(k+l),t} \end{bmatrix}$$

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

or

$$\Delta x_t = \gamma x_{t-1} + \sum_{2}^{p} \lambda_i \Delta x_{t-i+1} + \varepsilon_t$$

where $\Delta$is the differencing operator, such that $\Delta x_t = x_t - x_{t-1}$ and x_t is a $(k+l) \times 1$ vector. It has an equivalent VAR (p) representation as described in the preceding section.

$$x_t = a_0 + \sum_{i=1}^{p} (I_{(k+l)} + \gamma + \lambda_i) x_{t-1} + \sum_{i=2}^{p-1} (\lambda_i - \lambda_{i-1}) x_{t-i} - \lambda_{p-1} x_{t-p} + \varepsilon_t$$

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price transmission by time series\codes\R code piece (CPI 25 goods)\VECM.r

11. Johansen co-integration test based on VAR and VECM

The Johansen (1988) procedure relies on the relationship between the rank of the matrix and its characteristic roots. Johansen suggests starting with a traditional vector auto regression model (VAR), select appropriate number of lags (p) based on AIC in the VAR model, and then estimate the vector error correction model (VECM) and determine the rank of the matrix of parameters. The co-integration of the system is tested using the maximum likelihood $L_{max}\,(r)$ which is a function of the co-integration rank r. Johansen describes two test methods: Trace Test and Maximum Eigenvalue Test.

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price transmission by time series\codes\R code piece (CPI 25 goods)\VECM.r

12. Granger- Causality test based on VAR

For the Granger- causality tests the vector of variables y_t is split into two sub-vectors x_t^1, x_t^2 with dimension *k* and *l*. For the rewritten 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 $a_{ij,h} = 0\ for\ h = 1,2,\ldots,p$. 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.

Note that the question of causal ordering of relationships among the variables in a VAR is not resolved in a VAR. That is, by definition, the VAR allows for bidirectional, intertemporal effects. By definition, if y_{t-i} is found to have a significant coefficient as a regressor conditioning x_t, then we can interpret this as indicating that y_{t-i} intertemporally causes x_t. In contrast, if y_t were found to be a significant regressor conditioning x_t, we cannot rule out the hypothesis that x_t contemporarily causes y_t. Based on this logic, we can examine intertemportal causal ordering within the VAR framework.

The corresponding R code could be found here:\Dropbox\TA\2012-VolatilityTransmission\CPI price transmission by time series\codes\R code piece (CPI 25 goods)\VAR.r

13.Short Run (SR) and Long Run (LR) impact analysis

To analyze SR impact, we can use both first difference VAR (given the first difference is stationary) and structural VECM (SVECM). To analyze LR impact, we can also use SVECM. Since the VAR has been presented above, we will focus on SVECM in the following context.

A just identified SVEC needs K(K-1)/2 restriction, where K is the number of variables. [*Note:* The concept of just identified/ over-identified/ under-identified is different from the concept of identification in the cross-sectional data analysis.] Then to test a particular unrestricted coefficient being zero, whether it is in the LR or the SR matrix, we simply set it up to zero and re-estimate this new over-identified SVEC. Then the test statistic follows $\chi^2(1)$ under Ho. If Ho is rejected, we conclude a SR or LR impact. And such test is called overidentification test.

14. Impulse Response Function

(1). Motivation:

With a VAR model or a VECM, we can investigate the short-run and long-run impact relation between variables in a system. However, the dynamic structure is not explicitly characterized. In order to investigate the dynamic effects of one or multiple shocks on a variable in the system and thereafter the whole system, we can use impulse response analysis.

(2). Definition:

For a K-dimensional stationary VAR (p) model,

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + U_t$$

By Wald decomposition, we can rewrite it as an vector moving average process, such that

$$Y_t = (I - A_1 B - \cdots - A_P B)^{-1} U_t = \Phi(B) U_t = \sum_{i=0}^{\infty} \phi_i U_{t-i}$$

where$cov(U_t) = \Sigma$ and Φ_i is the MA coefficients measuring the **impulse response function** (IRF). More specially, $\phi_{jk,i}$ represents the response of variable ***j*** to a unit impulse in variable ***k*** occurring ***i-th*** period ago. Variable k does not cause variable j if and only if $\phi_{jk,i} = 0$, i = 1, 2, · · ·."IRF are used to evaluate the effectiveness of a policy change, say increasing rediscount rate." (Jinlung Lin, 2006).

For non-diagonalΣ, we can transform the covariance matrix into a diagonal matrix to check one variable shock. For Σis a positive and symmetric matrix, by Cholesky decomposition, there exist a lower triangular matrix *P* such that$\Sigma = PP'$. Then let $\theta_i = \phi_i P,\ \ w_t = P^{-1} U_t$ and $E(w_t w_t') = I.$

Then Y_t can be rewritten as

$$Y_t = \sum_{i=0}^{\infty} \theta_i w_{t-i}$$

Let *D* be a diagonal matrix with same diagonals with$P$ and $W = PD^{-1}, \Lambda = DD'$. Then let $B_0 = I_k - W^{-1}, W = PD^{-1}, B_i = W^{-1} A_i$ and rewrite the original VAR(p) model as

$$Y_t = B_0 Y_{t-1} + \cdots + B_p Y_{t-p} + V_t$$

"Obviously, B_0 is a lower triangular matrix with all 0 diagonals. In other words, Cholesky decomposition imposes a recursive causal structure from the top variables to the bottom variables."

(3). IRF based on VECM

If there exists unit roots and co-integration, using VARs to estimate the impulse response would be inconsistent (Jinlung Lin, 2006). To achieve a consistent IRF estimates under unit root and co-integration, we need to use VECM.

To perform impulse response analysis, we need to first estimate a VECM using LR test or Trace test. Then convert the VECM back to VAR model. And follow the same procedure as in VAR thereafter.

15. 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,

$$y_t = \mathrm{x}_t^{'}\beta + \mathrm{z}_t^{'}\delta_j + \epsilon_t \;\; t = T_{j-1} + 1 \ldots . T_j$$

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

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

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 $\left(\widehat{T_1}, \ldots \widehat{T_m}\right)$

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

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

$$(\widehat{T_1}, \ldots \widehat{T_m}) = argmin_{(T_1,\ldots T_m)} S_T(T_1 \ldots T_m) = argmin_{(T_1,\ldots T_m)} W_T(T_1 \ldots T_m)$$

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.