“Efficient Semiparametric Detection of Changes in Trend”

Chuan Goh
Department of Economics
University of Toronto
June 2008
Trend-stationary processes:

- This paper is concerned with models of the form

\[ Y_t = d \left( \frac{t}{T} \right)^\top \gamma \left( \frac{t}{T} \right) + u_t, \ t = 1, \ldots, T \]

where \( d \left( \frac{t}{T} \right)^\top \gamma \left( \frac{t}{T} \right) \) is a deterministic trend function and \( u_t \) belongs to a mean-zero stationary process

- \( d \left( \frac{t}{T} \right) \equiv (d_1 \left( \frac{t}{T} \right), \ldots, d_k \left( \frac{t}{T} \right))^\top \) is a known function; its individual components are taken to be piecewise-Lipschitz continuous

- \( \gamma \left( \frac{t}{T} \right) \equiv (\gamma_1 \left( \frac{t}{T} \right), \ldots, \gamma_k \left( \frac{t}{T} \right))^\top \) is unknown
Statistical question of interest:

- Determining the adequacy of the model with $\gamma_j \left( \frac{t}{T} \right) \equiv \gamma_{j0}$, i.e., the adequacy of the model

$$Y_t = \sum_{j=1}^{k} d_j \left( \frac{t}{T} \right) \gamma_{j0} + u_t,$$

where each term $d_j \left( \frac{t}{T} \right) \gamma_{j0}$ in the trend function evolves smoothly as a function of time between possible successive breakpoints in $d_j(\cdot)$

- This involves essentially the development of a specification test for *segmented trend stationarity*
Trend- vs. difference-stationarity:

- Study of macroeconomic fluctuations—*are business cycles characterized as stationary fluctuations about a smooth trend function?*

  - In this case, one-off shocks do not have permanent effects on trend

- Challenge of esp. Nelson & Plosser (1982)—the dynamic behaviour of most historical U.S. macroeconomic aggregates is better described by an $I(1)$ process than by trend-stationarity

  - Difference-stationarity: Random shocks have permanent effects on trend; can think of trend changing unpredictably each period
Possible reconciliation: *Segmented* trend stationarity:

- Perron (1989), Rappoport & Reichlin (1989)—alleged unit-root behaviour in time series considered by Nelson & Plosser (1982) disappears when infrequent breaks are built into the trend function, i.e., *segmented trend stationarity* is a better model than difference-stationarity
Parameter of interest in this paper:

- Vector $\chi(\gamma)$ whose components are the *total variation norms* of the corresponding components in the trend parameter $\gamma$:

$$\chi(\gamma) = \left( \int_0^1 |\gamma_1'(s)| \, ds, \ldots, \int_0^1 |\gamma_k'(s)| \, ds \right)^\top$$

- Unit-root behaviour in $\{Y_t\}$ implies unpredictable changes each period in one or more components of $\gamma$, which translates into a value of

$$\sum_{j=1}^k \int_0^1 |\gamma_j'(s)| \, ds = \infty$$
Parameter of interest in this paper (cont’d):

- A finite number of discrete breaks or smooth continuous change in one or more components of \( \gamma \) leads to

\[
0 < \sum_{j=1}^{k} \int_{0}^{1} \left| \gamma_j'(s) \right| ds < \infty
\]

- How to test \( H_0 : \chi(\gamma) = 0 \)?
Focus on developing a test for \( H_0 : \nu_k^\top \chi(\gamma) = 0 \) based on observations \((Y_1, \ldots, Y_T)\):

- Seek a test that attains an appropriately defined semiparametric power envelope against contiguous alternatives to \( H_0 \):
  
  - The result will be a procedure with maximal power in the absence of strong assumptions regarding the form of allowable time variation in \( \gamma \left( \frac{t}{T} \right) \) and the nature of the data-generating process for the errors \( \{u_t\} \)
Exploit fairly well-known theory for semiparametric inference:

- Setting is that of inference regarding a *pathwise-differentiable parameter* $\kappa(\theta_0)$ in the context of a *regular statistical model*

- Regularity of a statistical model implies a LAN structure for parametric submodels centred at $\theta = \theta_0$
Efficiency bound for regular estimators:

- Local LAN structure for parametric submodels passing through $\theta = \theta_0$ implies the existence of a well-defined asymptotic variance bound for regular estimators, i.e., the existence of a meaningful Convolution Theorem for the limiting distribution of such estimators.

- Estimator regularity is related to stability of the asymptotic distribution in neighbourhoods of the parameter of interest, i.e.,

$$
\mathcal{L} \left( \sqrt{n} \left( \kappa_n - \kappa \left( \theta_0 + \frac{1}{\sqrt{n}} \delta \right) \right) \right) \bigg| \mathcal{P}_{\theta_0 + \frac{1}{\sqrt{n}} \delta}^n \rightarrow Q_0
$$

as $n \rightarrow \infty$ for all directions $\delta$. 
Regular estimators (cont’d):

- In such a setting, a regular estimator with minimal asymptotic variance will be *asymptotically linear* with an *influence function* equal to the *efficient influence function* $\psi_{\theta_0}$, i.e.,

  $$\kappa_n = \kappa(\theta_0) + \frac{1}{n} \sum_{i=1}^{n} \psi_{\theta_0}(X_i) + o_p \left( n^{-\frac{1}{2}} \right)$$

  as $n \to \infty$; where $\left| \psi_{\theta_0}(\cdot) \right| \in L_2(P_{\theta_0})$; $\int \psi_{\theta_0} dP_{\theta_0} = 0$.

- Clearly such estimators will be asymptotically normal; the asymptotic variance bound is given by $E \left[ \psi_{\theta_0} \psi_{\theta_0}^\top \right]$. 

11
The Convolution Theorem implies a notion of asymptotic optimality for tests regarding a scalar-valued parameter:

- An asymptotically locally UMP test of $H_0 : \kappa(\theta_0) = 0$ against $H_1 : \kappa(\theta_0) > 0$ can be constructed from an efficient estimator $\kappa_n$ of $\kappa(\theta_0)$:
  
  - This test will be of “Wald type”
  
  - Suppose $\tau_n^2 \xrightarrow{p} E[\psi^2_{\theta_0}]$, where $\psi_{\theta_0}$ is the efficient influence function

  - Then a test that rejects $H_0$ whenever $\frac{\sqrt{n}\kappa_n}{\tau_n}$ is large will be asymptotically most powerful against alternatives of the form

$$H_{1n} : \kappa(\theta_0) + \frac{1}{\sqrt{n}}\dot{\kappa}_{\theta_0}(\delta)$$

for all directions $\dot{\kappa}_{\theta_0}(\delta), \delta \in \dot{\Theta}$
Back to model of interest:

\[ Y_t = d \left( \frac{t}{T} \right) ^\top \gamma \left( \frac{t}{T} \right) + u_t, \; t = 1, \ldots, T \]

- Stochastic component: \( u_t \) strictly stationary and ergodic, mean zero, finite variance

- Model parameter: \( \theta \equiv \left( \gamma, G(t), F(t|t-1) \right) \)
  - \( G(t) \) the joint distribution of \(( \ldots, u_{t-1}, u_t)\)
  - \( F(t|t-1) \) the conditional distribution of \( u_t \) given \( \sigma (\ldots, u_{t-1}, u_t)\)

- Want an asymptotically locally UMP test of \( H_0 : \nu_k ^\top \chi(\gamma_0) = 0 \)

under the sequence of distributions \( \left\{ P(\gamma_0, G(0), F_0(T|T-1)) : T \geq 1 \right\} \)
Derivation of efficient influence function:

- Impose conditions on $F_0^{(t|t-1)}$ that guarantee regularity of model
  \[
  \left\{ P \left( \gamma_0, G_0^{(T)}, F_0^{(T|T-1)} \right) : T \geq 1 \right\}
  \]

  - In particular, assume the Fisher information for location $J$ of $F_0^{(t|t-1)}$ is finite

  - Then the tangent $\dot{P}_{\theta_0}[\delta]$ of $P_\theta$ at $\theta_0$ has the form
    \[
    \dot{P}_{\theta_0}[\delta] = d(s)^\top a(s) l(u) + b(u_0) + c(u),
    \]
    where $\delta \equiv (a(\cdot), b(\cdot), c(\cdot)) \in \dot{\Theta}$, $l \equiv -\frac{f_0^{(t|t-1)'}}{f_0^{(t|t-1)}} 1 \{ f_0^{(t|t-1)} > 0 \}$ is the score of $F_0^{(t|t-1)}$. 

14
Derivation of efficient influence function (cont’d):

- The pathwise derivative of the parameter of interest \( \kappa(\theta) \equiv \chi(\gamma) \) along \( \dot{\Theta} \) at \( \theta = \theta_0 \) has the form

  \[
  \dot{\kappa}_{\theta_0}(\delta) = \dot{\chi}_{\gamma_0} \equiv \left( \int_0^1 |\gamma'_{01}(s)|a_1(s)ds, \ldots, \int_0^1 |\gamma'_{0k}(s)|a_k(s)ds \right)^T
  \]

- Can solve for the efficient influence function \( \psi_{\theta_0} \), which satisfies

  \[
  \int \psi_{\theta_0} \dot{P}_{\theta_0}[\delta]dP_{\theta_0} = \dot{\kappa}_{\theta_0}(\delta)
  \]
Efficient influence function (cont’d):

- In particular, $\psi_{\theta_0} = v_0(s)\frac{l(u_t)}{J} + \bar{v}u_t$, where for $s \in (0, 1)$ and $u_t \sim F_0^{(t|t-1)}$,

$$
\begin{align*}
    v(s) &\equiv (|\gamma'_{01}(s)|, \ldots, |\gamma'_{0k}(s)|)^\top \\
    \bar{v} &\equiv \left(\int_0^1 |\gamma'_{01}(s)| ds, \ldots, \int_0^1 |\gamma'_{0k} g s| ds\right)^\top \\
    v_0(s) &\equiv v(s) - \bar{v} \\
    l(u_t) &\equiv -\frac{f_0^{(t|t-1)'}}{f_0^{(t|t-1)}} 1\{f_0^{(t|t-1)}>0\} \\
    J &\equiv \int l^2 dF_0^{(t|t-1)}.
\end{align*}
$$
Construction of an efficient test of $H_0 : \chi(\gamma_0) = 0$:

- Find a suitable estimate $\hat{\psi}_{t,T}$ of $\psi_{\theta_0}$. Use $\hat{\psi}_{t,T}$ to construct the estimate

$$\hat{\chi}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{\psi}_{t,T}$$

of $\chi(\gamma_0)$.

- $\hat{\chi}_T$ will be efficient if

$$\hat{\chi}_T = \frac{1}{T} \sum_{t=1}^{T} \psi_{\theta_0} + o_p \left(T^{-\frac{1}{2}}\right).$$
Ingredients of an estimate of the efficient influence function:

- Conceptually simple plug-in estimate of the efficient influence function will be used

- Major ingredients are “Priestley-Chao” estimators of the trend parameter $\gamma$ and its derivative $\gamma'$:

  \[
  \hat{\gamma}_T(s) \equiv \left( d(s)d(s)\top \right)^{-1} d(s) \sum_{t=1}^{T} w_{l,t,T}(s)Y_t,
  \]

  where $w_{l,t,T}(s) \equiv \frac{1}{Th_l}K\left(\frac{1}{h_l}\left(\frac{t}{T} - s\right)\right)$, $h_l \to 0$, $Th_l \to \infty$ as $T \to \infty$; and

  \[
  \hat{\gamma}'_T(s) \equiv \frac{d}{ds}\hat{\gamma}_T(s)
  \]

  but with a different bandwidth $h_d$ with $Th_d^3 \to \infty$ as $T \to \infty$
Other ingredients:

\[\hat{u}_{t,T} \equiv Y_t - d \left( \frac{t}{T} \right) \tilde{\gamma}_T \left( \frac{t}{T} \right);\]
\[\hat{v}_T(s) \equiv \left( |\tilde{\gamma}'_{T1}(s)|, \ldots, |\tilde{\gamma}'_{Tk}(s)| \right)^\top;\]
\[\bar{v}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{v}_T \left( \frac{t}{T} \right);\]
\[\delta_{t,T} \equiv d \left( \frac{t}{T} \right)^\top \left( \tilde{\gamma}_T \left( \frac{t}{T} \right) - \gamma \left( \frac{t}{T} \right) \right);\]
\[\hat{f}_{UT}(u) \equiv \frac{1}{T} \sum_{t=1}^{T} \frac{1}{a_T} k \left( u - \hat{u}_{t,T} \right);\]
\[\hat{l}_{UT}(u) \equiv - \frac{\hat{f}'_{UT}(u)}{\hat{f}_{UT}(u) + b_T};\]
\[\hat{J}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \hat{l}_{UT}(\hat{u}_{t,T});\]

where \(0 < a_T, b_T \to 0\) as \(T \to \infty\) and \(k(\cdot)\) is another kernel function.
Efficient estimate of $\chi(\gamma_0)$:

$$\hat{\chi}_T \equiv \frac{1}{T} \sum_{t=1}^{T} \left[ \hat{v}_T \left( \frac{t}{T} \right) + J_T^{-1} \left( \hat{v}_T \left( \frac{t}{T} \right) - \bar{v}_T \right) \hat{l}_{UT} \left( \hat{u}_{t,T} \right) + \bar{v}_T \hat{u}_{t,T} \right].$$

Conditions on bandwidths $h_l, h_d, a_T, b_T$ and kernels $K(\cdot), k(\cdot)$ can be given that yield

$$\hat{\chi}_T = \frac{1}{T} \sum_{t=1}^{T} \left[ v \left( \frac{t}{T} \right) + J^{-1} \left( v \left( \frac{t}{T} \right) - \bar{v} \right) l_U(u_t) + \bar{v} u_t \right] + o_p \left( T^{-\frac{1}{2}} \right),$$

in regions of the parameter space shrinking to $\theta_0$ at rate $T^{-\frac{1}{2}}$ when the regression errors $\{u_t\}$ satisfy a general short-range dependence condition.
Strict stationarity, ergodicity and short-range dependence of \( \{u_t\} \):

1. \( u_t = H(\ldots, \epsilon_{t-1}, \epsilon_t) \) for an iid process \( \{\epsilon_t\} \), a measurable function \( H(\cdot) \) that makes \( u_t \) a random variable with mean zero and finite variance

2. \( E[|u_t|^4] < \infty. \)

3. For an iid copy \( \epsilon'_t \) of \( \epsilon_t \), define

   \[
   u_t^* \equiv H(\ldots, \epsilon_{-1}, \epsilon'_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t).
   \]

   Then

   \[
   \sum_{t=1}^{\infty} t \left( E[|u_t - u_t^*|^4] \right)^{\frac{1}{4}} < \infty
   \]
Strong invariance principle of Wu (2007):

- Wu & Shao (2004): The condition $|u_t - u_t^*| = O(r^T)$ ($r \in (0, 1)$) holds for many popular error processes, i.e., the short-range dependence condition is widely applicable.

- In these cases, the strong invariance principle of Wu (2007) is applicable:

$$\max_{t \leq T} \left| \sum_{s=1}^{t} u_s - \sigma B(t) \right| = o_{a.s.} \left( T^\frac{1}{4} \log T \right),$$

where $\sigma^2 \equiv \sum_{t=-\infty}^{\infty} E [u_0 u_t]$, $B$ is standard Brownian motion.
Suppose $\{Z_t\}$ is iid $N(0, \sigma^2)$:

- Under the general short-range dependence condition, weighted sums $\sum_{t=1}^{T} \omega_{t,T} u_t$ (where $\omega_{t,T}$ is of bounded variation) behave asymptotically like $\sum_{t=1}^{T} \omega_{t,T} Z_t$

- Therefore the large-sample theory under the short-range dependence of estimators involving kernel smoothing of $u_t$ may be mimicked by considering the behaviour of the same estimators applied to the model

$$Y_t = d \left( \frac{t}{T} \right)^\top \gamma \left( \frac{t}{T} \right) + Z_t$$

with iid $N(0, \sigma^2)$ errors

- Use this idea to analyze the behaviour of $\hat{\chi}_T$ under $P\left(\gamma_0, G_0^{(T)}, F_0^{(T|T-1)}\right)$
Asymptotic normality of efficient estimator:

- Central result of paper: Under certain conditions on \( \{u_t\} \) and \( F_0^{(t|t-1)} \), bandwidths \( h_l, h_d, a_T \) and \( b_T \), and kernels \( K(\cdot) \) and \( k(\cdot) \) get that

\[
\frac{\sqrt{T} \mathbf{v}_k^\top \hat{\chi}_T}{\sqrt{\mathbf{v}_k^\top \hat{\Psi}_T \mathbf{v}_k}},
\]

where

\[
\hat{\Psi}_T \equiv \frac{1}{T} \sum_{t=1}^T \hat{\psi}_{t,T} \hat{\psi}_{t,T}^\top,
\]

is asymptotically \( N(0, 1) \) under \( H_0 : \chi(\gamma_0) = 0 \).
Efficient Wald-type test for parameter stability:

- A test of $H_0 : \chi(\gamma_0) = 0$ that rejects whenever

$$\frac{\sqrt{T \nu_k^T \hat{\chi}_T}}{\sqrt{\nu_k^T \hat{\Psi}_T \nu_k}} > z_{1-\alpha}$$

will be asymptotically locally UMP level-$\alpha$. 
Monte Carlo:

- Draw simulated data from

\[
\begin{align*}
Y_t &= \gamma\left(\frac{t}{T}\right) + u_t \\
u_t &= \rho u_{t-1} + \epsilon_t,
\end{align*}
\]

\(t = 1, \ldots, T; \quad \{\epsilon_t\} \text{ is iid } N(0, 1); \text{ 1000 replications}\)

- Test \(H_0: \gamma\left(\frac{t}{T}\right) \equiv \gamma_0\)

- Kernel \(k(\cdot)\): logistic density

- Standard normal kernel used to construct Priestley-Chao estimates of \(\gamma(\cdot)\) and its derivative
Empirical size performance at a nominal level of 5%:

- In order to examine size performance, simulate with $\gamma \left( \frac{t}{T} \right) \equiv 0$ and experiment with $T \in \{100, 200, 300\}; \rho = \frac{1}{2}$ and different bandwidth settings

- Take $h_{lT} = h_{dT} = h_T \equiv chT^{-\frac{2}{5}}; a_T = caT^{-\frac{2}{9}}; b_T = cbT^{-\frac{2}{9}}$; leading constants range over set $\{0.5, 1.0, 1.5\}$

- Rates of convergence of $h_T, a_T, b_T$ obtained by “asymptotic undersmoothing” with respect to the MISE-optimal rates for estimating a regression function ($h_T \propto T^{-\frac{1}{5}}$) and of a density derivative ($a_T \propto T^{-\frac{1}{9}}$)
Empirical size performance (cont’d):

- Accuracy of the first-order distributional approximation highly sensitive to bandwidths used

- The setting $c_h = c_a = c_b = 0.5$ appears to be fairly accurate
Empirical size performance at a nominal level of 5%:

<table>
<thead>
<tr>
<th>Bandwidth $c_h$ $c_a$ $c_b$</th>
<th>Sample size $T = 100$</th>
<th>$T = 200$</th>
<th>$T = 300$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 0.5 0.5</td>
<td>0.076</td>
<td>0.043</td>
<td>0.062</td>
</tr>
<tr>
<td>1.0 1.0 1.0</td>
<td>0.176</td>
<td>0.085</td>
<td>0.022</td>
</tr>
<tr>
<td>1.5 1.5 1.5</td>
<td>0.014</td>
<td>0.167</td>
<td>0.036</td>
</tr>
<tr>
<td>1.0 0.5 0.5</td>
<td>0.169</td>
<td>0.054</td>
<td>0.010</td>
</tr>
<tr>
<td>0.5 1.0 0.5</td>
<td>0.099</td>
<td>0.094</td>
<td>0.068</td>
</tr>
<tr>
<td>0.5 0.5 1.0</td>
<td>0.051</td>
<td>0.138</td>
<td>0.027</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Local power analysis under same simulated DGP:

• Take bandwidth constants \( c_h = c_a = c_b = 0.5 \), same rates of convergence of \( h_T, a_T, b_T \) as previous experiment; same kernels; \( T = 200 \) and 1000 replications

• Consider power successively against single and double breaks, continuous increase in mean, and a stable trend but a unit root in \( \{u_t\} \):

\[
\gamma\left(\frac{t}{T}\right) = \begin{cases} 
1 & \left(\frac{t}{T} > .5\right) \gamma_0, \\
\frac{1}{2} \times 1 & \left(\frac{25}{T} < \frac{t}{T} < .75\right) + 1 & \left(\frac{t}{T} \geq .75\right) \gamma_0, \\
1 \left(\frac{.25}{T} < \frac{t}{T} < .75\right) & \left[2 \left(\frac{t}{T} - \frac{1}{4}\right)\right] + 1 & \left(\frac{t}{T} \geq .75\right) \gamma_0, 
\end{cases}
\]

where \( \gamma_0 \) increases from 0 to 1; and also

\[
Y_t = \rho_0 u_{t-1} + \epsilon_t,
\]

where \( \rho_0 \) increases from 0.5 to 1.
Empirical power of a 5%-test against a one-time break in trend:
Empirical power of a 5%-test against two breaks in trend:
Empirical power of a 5%-test against a continuous linear change in mean:
Empirical power of a 5%-test against increasing persistence:

- Test adequacy of a linear trend-stationary model with a single break in level just after 1929, i.e., the model

\[ Y_t = \gamma_0 + d_2 \left( \frac{t}{T} \right) \gamma_0 + \frac{t}{T} \gamma_0 + u_t, \]

where \( d_2 (\cdot) \) is the indicator for observations corresponding to 1930 or later.

- Use standard normal kernel to construct the estimates of \( \gamma(\cdot) \) and \( \gamma'(\cdot) \)

- Try different bandwidths: \( h_l = c_l T^{-\frac{2}{5}} \) (\( c_l \in \{.125, 1.0, 8.0\} \)) for \( \hat{\gamma}_T \); \( h_d = c_d T^{-\frac{2}{7}} \) (\( c_d \in \{.125, 1.0, 8.0\} \)) for \( \hat{\gamma}'_T \)
Nonparametric fits of the Nelson & Plosser (1982) real GNP series:
Nonparametric fits of the first differences of the Nelson & Plosser (1982) real GNP series:

First differences of logarithm of real GNP, 1909–1970

- Actual series
- $c_d = 0.125$
- $c_d = 1.0$
- $c_d = 8$

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>Wald statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_l$</td>
<td>$c_d$</td>
</tr>
<tr>
<td>.125</td>
<td>.125</td>
</tr>
<tr>
<td>.125</td>
<td>1.0</td>
</tr>
<tr>
<td>.125</td>
<td>8.0</td>
</tr>
<tr>
<td>1.0</td>
<td>.125</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1.0</td>
<td>8.0</td>
</tr>
<tr>
<td>8.0</td>
<td>.125</td>
</tr>
<tr>
<td>8.0</td>
<td>1.0</td>
</tr>
<tr>
<td>8.0</td>
<td>8.0</td>
</tr>
</tbody>
</table>