On detecting end-of-sample instabilities

Fabio Busetti, Bank of Italy

Conference in honour of Andrew Harvey’s 65 year

Oxford, 29 June 2012
Motivation of the paper

- Failing to account for **structural changes** produces bias in the estimated coefficients of the econometric models and may result into poor (model-based) forecasts.

- Parameter instabilities that occur **towards the end of the sample** are particularly hard to detect (few observations). Commonly used tests of structural changes are not suited for this situation - low power.

- This paper considers modifications of existing tests and introduces new **statistics designed to have higher power against this hypothesis**. The focus is on the case of unknown changepoint.
Quick review of the literature on detecting instabilities (1)


3 Fluctuation and CUSUM-type tests. Brown-Durbin-Evans (1975), Ploberger-Kramer-Kontrus (1989), Ploberger-Kramer (1992). Without references to any specific alternative hypothesis (Harvey, 1975: "..... but this leaves us with a test which may be very weak when used in the presence of many types of structural change likely to occur in practice....").
3 **Fluctuation and CUSUM-type tests.** Brown-Durbin-Evans (1975), Ploberger-Kramer-Kontrus (1989), Ploberger-Kramer (1992). Without references to any specific alternative hypothesis (Harvey, 1975: "..... but this leaves us with a test which may be very weak when used in the presence of many types of structural change likely to occur in practice....").

What this paper does

- Build-up on the Andrews-Ploberger approach to design a **Wald-type statistic** that has higher power against end-of-sample structural changes.
What this paper does

- Build-up on the Andrews-Ploberger approach to design a **Wald-type statistic** that has higher power against end-of-sample structural changes

- Deliver new **LMP-type statistics** obtained assuming changes from stable to random walk coefficients at particular points in the sample → somehow generalize the test of Nyblom (1989). In a time series context, a similar approach was taken in Busetti-Taylor (2004).
What this paper does

- Buid-up on the Andrews-Ploberger approach to design a **Wald-type statistic** that has higher power against end-of-sample structural changes

- Deliver new **LMP-type statistics** obtained **assuming changes from stable to random walk coefficients at particular points in the sample** → somehow generalize the test of Nyblom (1989). In a time series context, a similar approach was taken in Busetti-Taylor (2004).

- Asymptotic distributions and critical values are provided. The power when there are **very few post-change observations** is compared with the subsampling-types tests of Andrews (2003) that assume a known changepoint.
What this paper does

- Build-up on the Andrews-Ploberger approach to design a **Wald-type statistic** that has higher power against end-of-sample structural changes.

- Deliver new **LMP-type statistics** obtained assuming changes from stable to random walk coefficients at particular points in the sample → somehow generalize the test of Nyblom (1989). In a time series context, a similar approach was taken in Busetti-Taylor (2004).

- Asymptotic distributions and critical values are provided. The power when there are **very few post-change observations** is compared with the subsampling-types tests of Andrews (2003) that assume a known changepoint.

- An **empirical illustration** of the use of the tests in the context of simple 'now-casting' models of Italian GDP and industrial production.
The framework

Linear regression model with $K = k_1 + k_2$ regressors $x_t = (x'_{1t}, x'_{2t})'$ and $T = n + m$ observations

$$y_t = \begin{cases} x'_{1t} \beta_1 + x'_{2t} \beta_2 + u_t & \text{for } t = 1, \ldots, n \\ x'_{1t} (\beta_1 + \delta_t) + x'_{2t} \beta_2 + u_t & \text{for } t = n + 1, \ldots, n + m \end{cases}$$

where $u_t$ is an i.i.d. disturbance

A changepoint may occur in the second part of the sample, of size $m$

The null hypothesis of parameter stability is

$$H_0 : \delta_t = 0 \quad \text{for all } t = n + 1, \ldots, n + m.$$ 

Under the alternative hypothesis $\delta_t \neq 0$ for some $t$. 

Fabio Busetti (Bank of Italy)  On detecting end-of-sample instabilities  Oxford, 29 June 2012  6 / 24
1a. Wald-type tests: known changepoint

The standard F-statistic against a one time shift in the coefficients of $x_{1t}$ at a known fraction $\pi = 1 - m / T \in (0, 1)$ of the sample is

$$F(\pi) = \frac{(Q - Q(\pi))/k_1}{Q(\pi)/(T - 2k_1 - k_2)}$$

where $Q$ denote the SSR under the restriction of constant coefficients and $Q(\pi)$ the SSR that allows a parameter change at $\pi$.

Under $H_0$, $F(\pi) \rightarrow k_1^{-1} \chi^2(k_1) \overset{d}{=} F_\infty(\pi) \equiv k_1^{-1} \frac{B_{k_1}(\pi)'B_{k_1}(\pi)}{\pi(1-\pi)}$, where $B_{k_1}(\cdot)$ is Brownian bridge process. This is a limiting result for $T \rightarrow \infty$; thus the finite sample approximation is good if $m$ is not too small.
If location of the shift is not known a priori, Quandt (1960) proposes to take the **maximum of the F-statistics** over the set of possible changepoints,

\[
    \text{Sup-F} = \sup_{\pi \in \Pi} F(\pi)
\]

where \(\Pi\) is a closed subset of \((0, 1)\). Under \(H_0\), \(\text{Sup-F} \xrightarrow{d} \sup_{\pi \in \Pi} F_\infty(\pi)\).

Andrews and Ploberger (1994): **taking averages** yield to somewhat better properties, e.g.

\[
    \text{Exp-F} = \log \int_{\pi \in \Pi} \exp \left( \frac{k_1}{2} F(\pi) \right) dJ(\pi),
\]

where \(J(\pi)\) is a chosen weight function (i.e. probability measure) on the values of \(\pi \in \Pi\). They provide c.v.'s for constant weights.
In applied works, the set of possible breakpoints $\Pi$ is usually taken as $[.15, .85]$ as in the original paper —→ low power if changepoints occur near the end of the sample. On the other hand, searching for $\pi$ close to 1 may deteriorate the null limiting approximation (size of the test).

Here we explore **two ways to increase power** against end-of-sample changepoints.

- A) Take $\Pi = [.05, .95], [.01, .99], [.75, 99], [.90, .99]$ for Sup-F and Exp-F
1c. Wald-type tests: end-of-sample case

In applied works, the set of possible breakpoints $\Pi$ is usually taken as $[.15, .85]$ as in the original paper —— $>$ low power if changepoints occur near the end of the sample. On the other hand, searching for $\pi$ close to 1 may deteriorate the null limiting approximation (size of the test).

Here we explore **two ways to increase power** against end-of-sample changepoints.

- A) Take $\Pi = [.05, .95], [.01, .99], [.75, 99], [.90, .99]$ for Sup-F and Exp-F
- B) For Exp-F consider increasing weights over $\Pi = [.01, .99]$  

\[
\text{Exp-F}_{LIN} = \log \int_{\pi \in \Pi} \exp \left( \frac{k_1}{2} F(\pi) \right) \pi \, d\pi
\]
The LMP statistic against a switch from constant to random walk coefficients at a known fraction \( \pi = 1 - m/T \in (0, 1) \) of the sample is

\[
L(\pi) = \hat{\sigma}^{-2} (T - [\pi T])^{-2} \sum_{t=[\pi T]+1}^{T} S_t' V^{-1} S_t
\]

where \( \hat{u}_t = y_t - x_t' \hat{\beta} \) are the OLS residuals from regressing \( y_t \) on \( x_t \), \( \hat{\sigma}^2 = (T - K)^{-1} \sum_{t=1}^{T} \hat{u}_t^2 \), \( S_t = \sum_{j=t}^{T} \hat{u}_j x_{1j} \) and \( V = T^{-1} \sum_{t=1}^{T} x_{1t} x_{1t}' \).

This generalizes the statistic of Nyblom (1989) that corresponds to \( \pi = 0 \). Under \( H_0 \), \( L(\pi) \xrightarrow{d} (1 - \pi)^{-2} \int_{\pi}^{1} B_{k_1}(s)' B_{k_1}(s) ds \), where \( B_{k_1}(.) \) is Brownian bridge process. This is a limiting result for \( T \to \infty \); thus the finite sample approximation is good if \( m \) is not too small.
For the case of unknown changepoint towards the end of the sample, we consider the statistics

\[
\text{Sup-}L = \sup_{\pi \in \Pi} L(\pi)
\]

\[
\text{Exp-}L = \log \int_{\pi \in \Pi} \exp(L(\pi)) \, dJ(\pi)
\]

for \( \Pi = [.05, .95] \), [.01, .99], [.75, 99] and [.90, .99].

And also a test that gives higher weight to changepoints occurring later

\[
\text{Exp-}L_{LIN} = \log \int_{\pi \in \Pi} \exp(L(\pi)) \, \pi \, d\pi.
\]

The asymptotic critical values for these tests are provided in the paper.
3. When the number of post-changepoint observations is 'small'

Andrews (2003) has proposed a variant of the $F$ test, where the critical values are obtained by 'parametric subsampling'. The statistic is defined as

$$S = S_{n+1} \left( \hat{\beta}, \hat{\sigma}^2 \right)$$

where $\hat{\beta}$ and $\hat{\sigma}^2$ are as usual (using all the $n + m$ observations), and for $j = 1, 2, ..., n + 1,$

$$S_j (\beta, \sigma^2) = \sigma^{-2} \left( Y_j(m) - X_j(m)\beta \right)' P_j(m) \left( Y_j(m) - X_j(m)\beta \right)' ,$$

with $P_j(m) = X_j(m) [X_j(m)'X_j(m)]^{-1} X_j(m)'$, $X_j(m)$ is the $m \times K$ matrix $\left( x'_j, x'_{j+1}, ..., x'_{j+m-1} \right)'$, $Y_j(m)$ is the $m \times 1$ vector $(y_j, y_{j+1}, ..., y_{j+m-1})'$. This quadratic form essentially corresponds to the numerator of $F$-statistics for testing stability over rolling subsamples of size $m$. 
Andrews (2003) shows that the distribution function of $S$ can be well approximated by the empirical distribution of

$$\{ S_j (\beta, \sigma^2) : j = 1, \ldots, n - m + 1 \}$$

evaluated at consistent estimators of $\beta$ and $\sigma^2$. The critical value is the $1 - \alpha$ sample quantile of $\{ S_j (\beta, \sigma^2) : j = 1, \ldots, n - m + 1 \}$.

A similar approach can be used for the $L(\pi)$ statistic: it just requires to obtain the empirical distribution function of

$$\{ L_j \left( \hat{\beta}_{(n)}, \hat{\sigma}^2_{(n)} \right) : j = 1, \ldots, n - m + 1 \},$$

where

$$L_j \left( \hat{\beta}_{(n)}, \hat{\sigma}^2_{(n)} \right) = \hat{\sigma}_{(n)}^{-2} m^{-2} \sum_{t=j}^{j+m-1} S_{t,j} \left( \hat{\beta}_{(n)} \right)' V^{-1} S_{t,j} \left( \hat{\beta}_{(n)} \right),$$

and

$$S_{t,j} \left( \hat{\beta}_{(n)} \right) = \sum_{h=t}^{j+m-1} \left( y_h - x_h' \hat{\beta}_{(n)} \right) x_{1h}.$$
First, the 'static' regression model

\[ y_t = \begin{cases} 
  x_t' \beta + u_t & \text{for } t = 1, \ldots, n \\
  x_t' (\beta + \delta_t) + u_t & \text{for } t = n+1, \ldots, n+m
\end{cases} \]

where \( x_t = (1, (-1)^t) \), \( u_t \) is iid(0,1), with sample size \( T = n + m = 100, 200, 400 \).

Under the local alternative hypothesis the changepoint occurs at \( \pi_0 = .75, .90, .95, .98 \). For \( t \geq n + 1 \), consider the two cases

A) one time shift: \( \delta_t = \mathbf{i} c / \sqrt{T} \) where \( \mathbf{i} = (1, 1)' \)

B) random walk coefficients: \( \delta_t = \mathbf{i} \sum_{j=n+1}^{t} \eta_j \), where \( \eta_t \sim Niid(0, q^2 / T^2) \)
Size of selected tests: \( T=200 \) and Gaussian errors

<table>
<thead>
<tr>
<th>Test</th>
<th>0.75</th>
<th>0.90</th>
<th>0.95</th>
<th>0.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>F ( ) - subsampling</td>
<td>0.17</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Exp-F [.05,.95]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Exp-F [.75,.99]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>Exp-F-LIN [.01,.99]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>L ( ) - subsampling</td>
<td>0.15</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Exp-L [.05,.95]</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Exp-L [.75,.99]</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>Exp-L-LIN [.01,.99]</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Summary of the size properties

- For the tests shown in the table, good size properties for $T=200$
- Some oversizing occurs for LMP-type tests, particularly if the set of possible changepoints is narrow: $\Pi = [.90, .99]$
- Except for the subsampling-based tests (with known changepoint), the size deteriorates for heavy tails and asymmetric error distributions (the worsening is larger for the Wald-type statistics)
Power of the tests against a one time coefficient shift at PI=0.95

- F-subsampling with known PI (Andrews, 2003)
- Exp-F [.05,.95] (Andrews-Ploberger, 1994)
- Exp-F [.75,.99]
- Exp-F LIN
- L-subsampling with known PI
- Exp-L [.75,.99]
- Exp-L LIN
Power of the tests against a random walk coefficients starting at $\pi = 0.95$

- F-subsampling with known PI (Andrews, 2003)
- Exp-F [.05,.95] (Andrews-Ploberger, 1994)
- Exp-F [.75,.99]
- Exp-F LIN
- L-subsampling with known PI
- Exp-L [.75,.99]
- Exp-L LIN
Summary of the power properties

- For known end-of-sample changepoint Wald and LMP test have similar properties under cases of one-time shift and random walk coefficients.
- If changepoint is unknown, power is significantly higher for LMP-tests.
- Exp-type tests more powerful than Sup-type tests (as in Andrews and Ploberger, 1994).
- Similar properties of $\text{Exp-}L_{LIN}$ and $\text{Exp-}L$ over $\Pi = [.75, .99]$, but the former is more powerful if changepoint occurs earlier in the sample.
- For dynamic models, the Wald (LMP) type tests are more powerful against a change toward a lower (higher) degree of persistence.
Empirical illustration on *now-casting* models for Italian output

The purpose is to detect possible end-of-sample instabilities in two simple *now-casting* regression models:

a) a monthly model for **Industrial Production** as a function of **Business Confidence**

b) a quarterly model for **GDP** as a function of **Industrial Production**

\[ \Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta x_t + \text{error} \]

As business confidence is a very timely indicator, the two models can be used in conjunction for nowcasting Italian GDP.
Main findings

- The relationship between IP and confidence indicator appears stable if it is estimated using data up to the first half of 2008. Thereafter nearly all tests strongly reject the null hypothesis. Similar evidence of Wald-type and the LMP-type tests is very similar.

On detecting end-of-sample instabilities
Oxford, 29 June 2012
Main findings

- The relationship between **IP and confidence indicator** appears stable if it is estimated using data up to the first half of 2008. Thereafter *nearly all tests strongly reject the null hypothesis*. Similar evidence of Wald-type and the LMP-type tests is very similar.

- As regards the relationships between **GDP and IP**, the Wald-type tests almost never reject the null hypothesis of stability while *the LMP-type tests display a strong tendency to reject*, particularly when the data of 2009 are included in the sample. Consistent with simulation results.

---

*Fabio Busetti (Bank of Italy)*

**On detecting end-of-sample instabilities**

Oxford, 29 June 2012
Main findings

- The relationship between IP and confidence indicator appears stable if it is estimated using data up to the first half of 2008. Thereafter nearly all tests strongly reject the null hypothesis. Similar evidence of Wald-type and the LMP-type tests is very similar.

- As regards the relationships between GDP and IP, the Wald-type tests almost never reject the null hypothesis of stability while the LMP-type tests display a strong tendency to reject, particularly when the data of 2009 are included in the sample. Consistent with simulation results.

- A general message is that model-based predictions should be interpreted with caution in the presence of unusually large fluctuations of the indicators towards the end of the sample (call for forecaster’s judgement).
Tests constructed in terms of **LMP-type statistics** appear in general more powerful than those based on Wald statistics; the latter however possess better finite sample size properties.
Concluding remarks

- Tests constructed in terms of **LMP-type statistics** appear in general more powerful than those based on Wald statistics; the latter however possess better finite sample size properties.

- Overall, a LMP-type test that gives increasing weight to possible changepoints along the sample appears to be a good option.
Concluding remarks

- Tests constructed in terms of **LMP-type statistics** appear in general more powerful than those based on Wald statistics; the latter however possess better finite sample size properties.

- Overall, a LMP-type test that gives increasing weight to possible changepoints along the sample appears to be a good option.

- Connection with literature on monitoring instabilities in real time (e.g. Chu, Stinchcombe and White, 1996): sequential application of the test eventually yields a rejection of $H_0$; to control size critical values should depend on monitoring time. Scope for further research.