STRUCTURAL CHANGE AND THE PROBLEM OF PHANTOM BREAK LOCATIONS

by YAO RAO and BRENDAN MCCABE
Management School, University of Liverpool

It is well known, in structural break problems, that it is much easier to detect the existence of a break in a data set than to determine the location of such a break in the sample span. This paper investigates why, in the context of Gaussian linear regressions, using a decision theory framework. The nub of the problem, even for moderately sized breaks, is that the posterior probability distribution of the possible break points is usually not very informative about the true break location. The information content is measured here by a proper scoring rule. Hence, even a locally optimal break location procedure, as introduced here, is ineffective. In the regression context, it turns out to be quite common, indeed the norm, for break location procedures to misidentify the true break position up to 100 per cent of the time. Unfortunately too, the magnitude of the difference between the misidentified and true break locations is usually not small.

1 INTRODUCTION

Structural break procedures typically proceed by first determining if a break exists in a sample and, if so, they suggest a location in the sample span for the break to have occurred. Unfortunately, even for breaks that are of moderate magnitude relative to a given sample size, it appears that detecting the existence of a break is fairly easy while subsequently identifying the correct location is very difficult and sometimes almost impossible. This leads to the conclusion that many suggested break locations are potentially misidentified because, as argued persuasively by Elliott and Muller (2007), small breaks appear to be empirically very relevant. The objective of this paper is to explore this issue of phantom break locations in the context of Gaussian linear regressions. We first derive a locally optimal procedure for finding a break location in a (Bayesian flavored) frequentist decision theory framework. This provides a baseline against which other procedures may be compared. In this way, we also directly analytically address the question of how difficult it is to determine the location of breaks by examining the (posterior) probability distribution of the possible break locations. The basic

* Manuscript received 29.5.18; final version received 24.5.19.
problem is that the posterior distribution of the possible break points will contain little information on the true break position for small to moderate break sizes. The information content of the break point distribution relative to the true break point is measured by a proper scoring rule; the spherical score. This analysis is valid in finite samples and does not rely on embedding a model in, a possibly artificial, sequence of experiments, as would be necessary for a large sample asymptotic analysis.

A locally optimal procedure is suggested and it is based on the posterior distribution of the break points which is approximated by a residual CUSUM process. The maximum of the cusum process tests for the existence of a break and the argmax identifies the location, when the test rejects. The procedure can easily reject the hypothesis of no break (power), but it is not able, despite the optimality, to overcome the potential lack of information in the sample and identify the correct break location reliably (an ability we might call sensitivity). The power of break procedures has been studied quite extensively (e.g. Deng and Perron, 2008; Xu, 2013; Forchini, 2012; Andrews et al., 1996; Ploberger and Krämer, 1990) but the sensitivity much less so.

Bai and Perron (1998) study the behavior of the estimated break fraction in linear regression models and show it is consistent. The analysis in these studies is asymptotic in the sample size. By contrast, in the present paper, we show local (in the break parameter) optimality for the estimator of the break location.

However, the CUSUM procedure can misidentify the true break point 100 per cent of the time for larger sample sizes as the argmax is shown to be a biased estimator of the true break point for a representative broken trend model. This makes misidentifying break locations endemic in this situation. Unfortunately, it is not the case that, when the CUSUM procedure misidentifies a break location, the magnitude of the error is small and, perhaps, of little practical significance. Large sensitivity errors are very frequent, especially at smaller sample sizes.

The structure of the paper is as follows. We set up the regression decision model in Section 2 and derive the locally optimal decision rule in Section 3. In Section 4, we discuss the posterior distribution for break points while Section 5 introduces the broken trend model as an exemplifier and assesses the information content of the posterior. Section 5.2 provides a theoretical investigation of the behavior of the argmax, using orders of magnitude based on the sample size and shows that the argmax of the CUSUM is systematically biased for the true break location. Section 5.3 introduces a weighted modification, the $W$-CUSUM, that eliminates bias. It turns out that this $W$-CUSUM is equivalent to the popular structural break test of Bai (1997). Section 6 conducts some simulations. These simulations assess the information in the posterior of the model as well as the power and sensitivity of the CUSUM and $W$-CUSUM procedures. The simulated behavior of the argmax of the CUSUM procedure is also discussed in this Section and...
it is noted that the mode of the argmax distribution is biased for the true break point, in general, in conformity with the theoretical results in Section 5.2. While the $W$-CUSUM eliminates the bias, unfortunately this is accompanied by a relatively larger spread and is insufficient to overcome the lack of information in the posterior. The final Section concludes and suggests an empirical strategy.

2 THE REGRESSION DECISION MODEL

The set of models we consider for the observations is the multiple regression

$$y = X\beta + \omega_q\delta + \epsilon$$
$$\epsilon \sim N(0, \sigma^2 I)$$

(1)

where $y$ is a $T \times 1$ vector of observations, $X$ is a $T \times k$ full rank matrix of variables that is conditioned on, $\beta$ is a vector of unknown coefficients and $\epsilon$ is a vector of independent normal disturbances with zero mean and variance $\sigma^2$. The independent variables are assumed to be ordered in some way like a trend or seasonal in time series applications or by a treatment effect in cross sections. The regressors may be random and follow a stochastic process, for example, they may be stationary or unit root processes; the essential feature is that their distribution does not depend on $\delta$, hence $X$ is ancillary and therefore conditioned upon. The form of the structural break is captured by $\omega_q\delta$ with $\omega_q$ being a fixed vector and $\delta$ a scalar which may be positive or negative indicating the magnitude of the break. For example,

$$\omega_q = \left(0, \ldots, 0, q-1, 1, \ldots, 1\right)'$$

would model a shift in the intercept after position $q$.

We use the notation $r = My$, $M = I - X\left(X'X\right)^{-1}X'$, $\bar{\delta}^2 = r'r / (T-k)$ and $c_q = \omega_q'M\omega_q$. The distribution of the maximal invariant (under the addition to $y$ of vectors of the form $X\alpha$ and multiplication of $y$ by scalars $c \neq 0$) may be written in terms of the residuals, $r$, of the regression of $y$ on $X$ as

$$f_q(r|\delta) = c \exp \left\{-\frac{1}{2} \delta^2 c_q \right\} \left(r'r\right)^{(T-k)/2} \exp \left\{-\frac{1}{2} u^2 \right\} \cosh (u\delta) (u^{T-k-1})du$$

(2)

where $g(\lambda) = \int_0^\infty \exp \left\{-\frac{1}{2} u^2 \right\} \cosh (u\lambda) u^{T-k-1}du$. Here and throughout $c$ is a generic normalizing constant. Unfortunately, $f_q$ is pretty intractable and so we approximate locally in terms of the parameter $\delta$. This is done in the following theorem using the previously defined notation.

**Theorem 1**: When terms of order $\delta^4$ and above are ignored, an approximate expression for the maximal invariant (2) corresponding to the model (1), is given by

Electronic copy available at: https://ssrn.com/abstract=3613948
The proof of this theorem is in the Appendix.

3 Locally Optimal Break Locations Rules

We wish to decide between, $H_0$, a model with no break or, $H_2$, a break at position $q = 2$ specified by $\omega_2 \delta$ or, $H_3$, a break at position $q = 3$ specified by $\omega_3 \delta$ and so on up to position $q = T$. Thus, we have a multiple decision problem with $T$ models to choose from. It follows from Ferguson (1961, 1967) that an optimal rule for deciding if a break exists and if so, where in the sample span it is located, takes the form

$$f_q^a (r|\delta^2) = c (r' r)^{-(T-k)} \left( 1 + \frac{1}{2} \delta^2 \left( \frac{(\omega_q' r)^2}{\hat{\sigma}^2} - c_q \right) \right). \quad (3)$$

Here $f_q$ is the fully specified density of the observations and each of the possible break points is assigned prior weights/probabilities $p_q$. The constant $K$ is used to determine the size of the initial test and the rule is uniform in $\delta^2$. Unfortunately, it is not possible to use $f_q$ in (2) as it is intractable. However, using the approximation $f_q^a$ in (3) allows an approximately optimal procedure to be constructed for any prior values $p_q$.

We now consider the scenario where we do not have any information favoring a break in one location over another and so the prior weights/probabilities of a break over the span are uniform. Using (3), the locally optimal procedure is based on the test

$$\text{Decide } H_0 : \text{ if } \max_{q \in [2, T]} p_q f_q < K$$
$$\text{otherwise} \quad (4)$$
$$\text{Decide } H_q^* : q^* \text{ being the argmax of } \max_q.$$

This procedure is locally invariant admissible, i.e. the probability of deciding a break at position $q$, given the break did occur at $q$, i.e. $P(\text{Dec } q|H_q^*)$, cannot be increased by any other invariant rule without decreasing the equivalent $P(\text{Dec } s|H_s)$ at some other point $s \neq q$, for small $\delta$. It is also a local Bayes rule when a $0 - 1$ loss function is used and no other rule with a smaller Bayes risk exists for small values of $\delta$. 

© 2019 The Author. The Manchester School published by The University of Manchester and John Wiley & Sons Ltd

Electronic copy available at: https://ssrn.com/abstract=3613948
4 Posterior Distribution for Break Points

Since we often have no real prior information on $\beta$ or $\sigma^2$ or the sign of $\delta$, we may consider the invariant $f_q$ as equivalent to likelihood of the data without the need to specify priors for these parameters in this decision problem. So, interpreting $f_q$ as the (approximate) likelihood and using a uniform prior on the breaks gives a posterior break distribution (mass function)

$$
\pi_q (\delta^2) \propto f_q (r|\delta^2) \propto \left(1 + \frac{1}{2} \delta^2 \left[ \frac{\left(\omega^2 r \right)}{\hat{\sigma}^2} - c_q \right] \right)
$$

(5)

where $\pi_q$ represents the posterior probability that a break occurs at position $q$. Hence, we see that the optimal rule uses the mode (most probable value) of the posterior distribution as the screening test. In addition, should the size of the break be small, the posterior will be quite flat (it may also be deduced that $df_q / d\delta|_{\delta=0} = 0$ in (2)) indicating there will be little precise information about the location of the break in the data. Furthermore, the posterior probability is spread across all $T$ (possibly large) break locations and, when combined with a flat profile (small $\delta$), it suggests that the probability allocated to any individual location $q$ may be small indeed.

To get a quantitative measure of the information in the posterior distribution we calculate a variant of a proper scoring rule as used in the evaluation of forecast density functions (see, for example, Gneiting et al., 2007). Scoring rules are loss functions where the action chosen is not represented a simple number but by an entire mass function (or density) and the loss incurred is a measure of the deviation between what the mass function predicted for outcomes and the actual realised future outcome itself. We use a variant of the spherical score i.e.

$$
S = \pi_s / \sigma_x
$$

where $\pi_s$ is the value of the posterior at the true breakpoint $s$ and $\sigma_x$ is the standard deviation of the $\pi_q$. Hence, $S$ rewards a posterior placing high weight on the true break location but penalises a large spread. Thus, a small value of $S$ indicates that a posterior will not be very informative about the location of the true break position relative to one with a large value. Of course, $S$ is only useful in theoretical investigations as $s$, the true break location, is not known in empirical contexts. The expression (7) in the Appendix is used for $f_q$ in (5) to compute $S$, as it is accurate to the order $\delta^6$. The expression is also used in Section 6 to assess the content of the posterior in simulations from the broken trend model, described next.
5 Broken Trend Model

A commonly used model is the broken (continuous) trend break given (in the usual time series notation) by

\[ y_t = \alpha + \beta t + \epsilon_t; t = 1, \ldots, T_B - 1 \]
\[ y_t = \alpha + \beta t + \delta (t - T_B) + \epsilon_t; t = T_B, \ldots, T \]  

which simultaneously introduces a change in slope and intercept. The matrix used to compute \( M \) is \( X = [1t] \), where \( 1 \) is a column of 1's and \( t = (1, 2, \ldots, T) \).

The breaks are specified by

\[ \omega_q = \begin{pmatrix} 0, \ldots, 0, 1, \ldots, (T - q + 1) \end{pmatrix} \]

while

\[ c_q = \omega_q' M \omega_q \quad q \in [2, T] \]

The test is then based on

\[ \max_{q \in [2, T]} \left\{ \frac{\left( \sum_{t=q}^{T} (t-q) r_t \right)^2}{\hat{\sigma}^2} - c_q \right\} \]

and hence the locally optimal procedure is essentially a mean corrected version of the ordinary CUSUM based on OLS residuals, see McCabe and Harrison (1980) and Ploberger and Krämer (1990, 1992).

5.1 The Information in the Posterior: Scoring Rule

We conducted a small simulation study to assess the information in the posterior of the broken trend model via the \( S \) statistic. We simulated the model (6) with \( \alpha = \beta = 1, \epsilon_i \text{iid } N(0, 10^2) \) and \( T = 100 \). The critical values were computed by the parametric bootstrap and 1000 replications were used. We use \( \mu = s/T \) to indicate the true break fraction. The magnitude of \( \delta = 0.6 \) was chosen as it is the smallest magnitude such that the power of the test was close to 1 for some value of \( \mu \) over the sample span. This strategy was chosen to help alleviate the trade-off between size of the break and the size of the sample.

In this simulation environment, we are able to calculate the \( S \) statistic for each replication since we know the true break location \( s \). Table 1 shows the average \( S \) over the 1000 replications while Tables 2 and 3 show the components \( \pi_s \) and \( \sigma_s \) for three values of \( \mu \). Table 3 indicates that the spread of the posterior of the break points is pretty constant over the span but decreases markedly as \( T \) gets larger. On the other hand from Table 2,
the weight allocated by the posterior to the true break point decreases as we move away from the center of the span. As $T$ increases the situation gets worse and less weight is allocated by $\pi_s$. Table 1 suggests that, overall, the shrinking spread is insufficient to offset the lack of precision in $\pi_s$.

The general conclusion of the Tables is that the approximate posterior does not accurately reflect the position of the true break except perhaps when it occurs at the center of the span. While these results refer specifically to the information in the posterior itself they are indicative of how the max and argmax of the posterior will behave. This suggests that the CUSUM procedure will perform relatively poorly in identifying break locations that occur away from the center of the span (in the broken trend model) and that the sensitivity will deteriorate there for larger sample sizes. The next Section investigates the behavior of the posterior process, using orders of magnitude in $T$, to shed further light on the behavior of the argmax of the CUSUM.

5.2 The Information in the Posterior: Analytic

We investigate the observed process

$$C_{q,s} = \frac{(\omega_q' r_s)^2}{\hat{\sigma}_s^2} - c_q$$

which is the posterior, up to a linear transform, and also the process on which the CUSUM procedure is based when the true breakpoint occurs at $s$ in (1). Specifically, we have

$$r_s = My_s = MX \beta + M \omega_s \delta + M \epsilon$$

$$= M \omega_s \delta + M \epsilon$$

and $\hat{\sigma}_s^2$ is the usual variance of the residuals. It does not seem feasible to establish the finite sample distribution of $\arg\max_q C_{q,s}$ But $E [C_{q,s}]$ is an
average or typical residual based profile over \( q \), given the break position \( s \). Thus, we suggest that arg \( \max_q E \left[ C_{q,s} \right] \) be investigated to assess the information in the posterior and to see how the identification procedure might behave when the true break point is at \( s \). Despite the fact that \( C_{q,s} \) involves ratios, we can approximate and use the ratio of the expectations i.e.

\[
E \left[ \frac{\left( \omega'_q r_s \right)^2}{\delta_s^2} \right] \approx \frac{E \left[ \left( \omega'_q r_s \right)^2 \right]}{E \left[ \delta_s^2 \right]}.
\]

In fact, when \( \delta = 0 \) we have a ratio of quadratic forms in normal variables and the ratio is independent of the denominator (See Pitman, 1937). The expectation of the ratio is then the ratio of the expectations and the approximation is exact. It may, therefore, also be considered accurate for small \( \delta \) and, in any event, the approximation is the first term in a Taylor series expansion for the ratio. The expectation in the numerator is

\[
E \left[ \left( \omega'_q r_s \right)^2 \right] = \left( \omega'_q M \omega_s \right)^2 \delta^2 + E \left[ \omega'_q M \varepsilon \varepsilon' M \omega_q \right]
\]

\[
\equiv c_{q,s} \delta^2 + \sigma^2 c_q.
\]

After some lengthy algebra outlined in the Appendix, we have, using \( \tau = q/T, \mu = s/T \) and the broken trend model,

\[
c_q = T^3 \frac{1}{3} \tau^3 (1 - \tau)^3 + o \left( T^3 \right)
\]

\[
c_{q,s} = \frac{1}{6} T^3 (1 - \max (\tau, \mu))^2 (3 |\mu - \tau| - 2 \max (\tau, \mu) + 2)
\]

\[
- \frac{1}{6} T^3 (1 - \tau)^2 (1 - \mu)^2 (\tau + \mu + 2 \tau \mu + 2) + o \left( T^3 \right)
\]

\[
\equiv \frac{1}{6} T^3 g (\tau, \mu) + o \left( T^3 \right)
\]

which gives the orders of magnitude as \( T^3 \) under the no break model. It also follows that

\[
E \left[ \frac{r'_s r_s}{T} \right] = \sigma^2 + \delta^2 T^2 \frac{1}{3} \mu^3 (1 - \mu)^3 + o \left( T^2 \right).
\]

Collecting terms, in orders of magnitude, we get

\[
E \left[ T^{-3} \frac{\left( \omega'_q r_s \right)^2}{\delta_s^2} \right] \approx \frac{T^3 \left( \frac{1}{6} g (\tau, \mu) \right)^2 \delta^2}{\sigma^2 + \delta^2 T^2 \frac{1}{3} \mu^3 (1 - \mu)^3} + \frac{\sigma^2 \tau^3 (1 - \tau)^3}{\sigma^2 + \delta^2 T^2 \mu^3 (1 - \mu)^3}
\]
which is of order $T$ when a break is present. Thus, our large $T$ approximate expression for calculating the argmax when $\delta \neq 0$ is

$$E \left[ T^{-4} \left( \frac{(\alpha_q r)^2}{\hat{\sigma}} - c_q \right) \right] \approx \left( \frac{1}{6} g(\tau, \mu) \right)^2 \frac{1}{\mu^3} (1 - \mu)^3$$

Even though this expression appears quite complex and $g$ is not differentiable in $\tau$, it is easy to derive the argmax by simply plotting the function over $\tau$. Given a true break point $\mu$, such a diagram shows (approximately) what a typical residual based profile looks like over $\tau$ as well as its argmax.

In Fig. 1, the profile of the CUSUM is the solid line shown in red. From the left panel, we see that the argmax of the profile does not match the true $\mu = 0.3$ and takes a value somewhat greater than 0.4. This over positioning by the CUSUM argmax is a general feature for $\mu < 0.5$. For $\mu = 0.5$ the argmax is properly positioned. For $\mu = 0.7$ the argmax understates the true break position. This too is a general feature for $\mu > 0.5$. We can surmise that outside of the center of the span (for the broken trend model), the argmax of the CUSUM distribution is systematically biased and hence the CUSUM procedure is incapable of finding the true break location reliably.

5.3 A Modified Procedure

It is interesting to enquire if the bias of the CUSUM could be removed by a suitable transformation. Now, $V \left[ \omega_q r \right] = \sigma^2 c_q$ and hence

![Fig. 1. Expected Residual Profile of the W-CUSUM and CUSUM](https://ssrn.com/abstract=3613948)
Thus the weighted cusum process has constant variance for all \( q \). This suggests the use of the \( W-CUSUM \)

\[
\max_{q \in [2, T]} c^{-1}_q \frac{(\omega'_q r)^2}{\hat{\sigma}^2} = \max_{q \in [2, T]} c^{-1}_q \frac{\sum_{t=q}^{T} (t-q) r_t^2}{\hat{\sigma}^2}
\]

in the broken trend model and it is equivalent to the minimum sum of squares test of Bai (1997).

Repeating the algebra of the previous section we get the expectation of the residual based profile

\[
E \left[ T^{-1} \frac{(\omega'_q r)^2}{c_q \hat{\sigma}^2} \right] \approx \frac{\left( \frac{1}{6} g(\tau, \mu) \right)^2}{\frac{1}{3} \mu^3 (1-\mu)^3 \frac{1}{3} \tau^3 (1-\tau)^3}.
\]

This is graphed in Fig. 1 and is shown as the dotted line in blue. The lack of bias in the argmax of the typical weighted residual based profile is evident. In the next Section, we conduct some simulation studies to get more detailed information on the behavior of the \( CUSUM \) and \( W-CUSUM \) processes.

6 Simulations

Using the same setup as in Section 5.1, we conducted a simulation to assess power and sensitivity of the \( CUSUM \) and \( W-CUSUM \) procedures. We also simulate the sampling distribution of the argmax itself. First, we discuss the \( CUSUM \) whose results are displayed in the right-hand panels of the diagrams below. Subsequently, we discuss the \( W-CUSUM \) whose results are in the left hand panels.

The right hand panel of Fig. 2 illustrates the performance of the \( CUSUM \) at \( T = 100 \); the top line shows the power (the ability to reject the no break model) while the lower line shows the sensitivity. It is clear that the test has greatest power in the center of the span but that tapers off toward the extremities. But, the most striking feature of the diagram is that, even when the power is close to 1, the procedure is not effective in identifying the exact location of the break (just 13 per cent of the time).

To give an idea of the spread of possible values chosen by the procedure we plot, in the right hand (\( CUSUM \)) panels of Fig. 3, the frequency distribution of the chosen break points (the argmax) for three true values of \( \mu \), again at \( T = 100 \). When the true \( \mu \) is 0.3, in the top panel, the mode of the argmax distribution badly overestimates the true break point, being located around
position 40. In the center panel, when the true position is 50, the modal value is well positioned but with a small positive bias while Panel three, where $\mu = 0.7$, is a mirror image of $\mu = 0.3$ with the mode of the argmax distribution underestimating the true breakpoint. These results are in conformity with the algebraic derivations of Section 5.2. In all cases, the argmax distribution based on the CUSUM has a quite large spread, increasing as $\mu$ deviates from 0.5. The implication is that the magnitude of the sensitivity error can be very large indeed in smaller sample sizes.

Next, the sample size was increased to $T = 250$, the smallest size such that the power was 1 for almost all of the span. The results are presented in Fig. 4 in the right-hand panel.

The ability to identify the correct location of the break improves at the $\mu = 0.5$ central position (from 13 per cent to 23 per cent) but in other areas of the span the sensitivity of the CUSUM procedure is essentially zero! The argmax frequency distributions are shown in Fig. 5 for various $\mu$ again on the right.

The variation of the argmax distributions shrinks but the biases do not. When the break deviates from the center of the span, there is no intersection whatsoever between the distribution and the true break position, for the larger sample size. In this situation, all the argmax suggestions of a break location are incorrect. The unfortunate implication of this is that while the procedure flags the existence of a break with probability close to 1, it almost certainly misidentifies the location, e.g. a phantom location is suggested 100 per cent of the time by the CUSUM over almost all of the span. Not only

FIG. 2. The Power and Percentage of Times the Correct Location was Identified for the W-CUSUM and CUSUM Procedures for $T = 100$
[Colour figure can be viewed at wileyonlinelibrary.com]
Fig. 3. ArgMax frequency distributions of the W-CUSUM and CUSUM for various $\mu$ and $T = 100$
[Colour figure can be viewed at wileyonlinelibrary.com]

© 2019 The Author. The Manchester School published by The University of Manchester and John Wiley & Sons Ltd
that but the size of the sensitivity error, while significantly reduced at the larger sample size, remains quite substantial. These findings are complimentary to the information gleaned from the scoring rule $S$ of Section 5.1.

The left-hand panels of Figs 2–5 show the performance of the $W$-CUSUM. From Fig. 2 we see that the power of the $W$-CUSUM is less than that of the CUSUM in the center of the span at $T = 100$. From Figs 2 and 4 we see the ability of the $W$-CUSUM to identify the correct break location is also less in the center of the span but marginally better near the extremities. From Fig. 3, it is clear that the bias has indeed been eliminated from the distribution of the $W$-CUSUM argmax but that the price paid is an increase in the spread of the distribution. This translates into much greater sensitivity error for the $W$-CUSUM. Figure 5 also demonstrates the lack of bias and the gain in precision for $T = 250$ but the ability to find the correct location is still quite low (under 10 per cent at best).

7 Conclusions

This paper investigated the phenomenon whereby structural change tests correctly recognize that a break has taken place but fail to correctly identify the location thereby giving rise to phantom break locations. The broken trend model and two CUSUM procedures are used to illustrate but the ideas are quite general. The basic problem is that the posterior distribution of the possible break points contains too little information on the true break position even for

Fig. 4. The Power and the Percentage of Times the Correct Location was Identified for the W-CUSUM and CUSUM Procedures for $T = 250$
[Colour figure can be viewed at wileyonlinelibrary.com]
Fig. 5: ArgMax Frequency Distributions of the W-CUSUM and CUSUM for Various $\mu$ and $T = 250$.

[Colour figure can be viewed at wileyonlinelibrary.com]

© 2019 The Author. The Manchester School published by The University of Manchester and John Wiley & Sons Ltd
moderately sized δ in relation to a given sample size. This lack of information is quantified by a variant of a spherical score measure. The two CUSUM procedures are compared and they both perform best in identifying the break location (though quite poorly!) if the break occurs at the center of the span for the broken trend model. The locally optimal CUSUM dominates the weighted version there for small δ. When a break occurs that is not in the center of the span, the performance of both procedures deteriorates with the CUSUM misidentifying the true break point 100 per cent of the time for larger sample sizes. An analysis based on orders of magnitude shows that the argmax of the CUSUM procedure is badly biased for the true break point outside the center of the span while that of the weighted version is not. Unfortunately removing the bias is insufficient to overcome the paucity of information in the posterior, resulting in the argmax distribution of the W-CUSUM having a very wide spread unless the sample size is large. In consequence, the magnitude of the difference between the misidentified and true break locations (sensitivity error) is much larger for the W-CUSUM than it is for the CUSUM in small and moderate samples.

These results suggest a strategy for applications. Apply both tests. If both tests accept the no break model, proceed on that basis. If one test only rejects, accept the recommended break location. If both tests reject, accept the recommendation of the W-CUSUM. Of course, this strategy cannot overcome lack of information about the break position in the sample. However, if there is some guidance as to the location of a possible break, induced by some major event say, then the observation span could be tailored so that the possible break location is near the center of the span (for the broken trend model) and/or non-uniform prior probabilities for the possible break locations employed. Other models may require positioning the potential break location at other fractions of the span but these may be deduced by the same methods that are employed here.

APPENDIX

Proof on Theorem

Since \( \cosh(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots \) and using \( x = \delta r_q, r_q = \omega / (r' r)^{1/2} \) we may approximate \( g(x) = \int_0^\infty \exp \left\{ -\frac{1}{2} u^2 \right\} \cosh (ux) u^{T-k-1} du \) to get

\[
\begin{align*}
\int_0^\infty \exp \left\{ -\frac{1}{2} u^2 \right\} \left[ 1 + \frac{u^2 x^2}{2!} + \frac{u^4 x^4}{4!} \right] u^{T-k-1} du \\
= \int_0^\infty \exp \left\{ -\frac{1}{2} u^2 \right\} u^{T-k-1} du + \frac{x^2}{2!} \int_0^\infty \exp \left\{ -\frac{1}{2} u^2 \right\} u^2 u^{T-k-1} du \\
+ \frac{x^4}{4!} \int_0^\infty \exp \left\{ -\frac{1}{2} u^2 \right\} u^4 u^{T-k-1} du \\
= k_1 + \delta^2 \frac{1}{2} k_2 r_q^2 + \delta^4 \frac{1}{4!} k_3 r_q^4
\end{align*}
\]
With a bit of algebra the integrals in the constants $k_1$ and $k_2$ may be computed using $\int_0^\infty \exp \left\{ -\frac{1}{2}u^2 \right\} u^m du = 2^{\frac{m+1}{2}}\sqrt{\pi} \Gamma \left( \frac{M+1}{2} \right)$ which gives $k_2 = k_1 (T-k)$, $k_3 = k_1 (T-k+2) / 2$ and $g (\delta r_q) \approx k_1 \left( 1 + \delta^2 \frac{1}{2} (T-k) r_q^2 + \delta^4 \frac{1}{48} (T-k+2) r_q^4 \right)

The maximal invariant $f_q = c \exp \left\{ -\frac{1}{2} \delta^2 c_q + \frac{1}{8} \delta^4 c_q^2 \right\} (r')^{- (T-k)} \left( 1 + \delta^2 \frac{1}{2} (T-k) r_q^2 + \delta^4 \frac{1}{48} (T-k+2) r_q^4 \right)

\begin{align*}
\quad
f_q &= c \left( 1 + \frac{1}{2} \delta^2 c_q + \frac{1}{8} \delta^4 c_q^2 \right) (r')^{- (T-k)} \left( 1 + \delta^2 \frac{1}{2} (T-k) r_q^2 + \delta^4 \frac{1}{48} (T-k+2) r_q^4 \right) \\
&= c \left( r' \right)^{- (T-k)} \left( 1 + \frac{1}{2} \delta^2 (T-k) r_q^2 - c_q \right) - \delta^4 \frac{1}{4} (T-k) c_q^2 r_q^2 + \delta^4 \frac{1}{48} (T-k+2) r_q^4 + \frac{1}{8} \delta^4 c_q^2
\end{align*}

\begin{align*}
\quad
f_q &= c \left( 1 + \frac{1}{2} \delta^2 \left[ (\omega_q')^2 \right] / \delta^2 - c_q \right) + \delta^4 \left( \frac{1}{4} c_q \delta^2 + \frac{1}{48} (\omega_q')^4 \delta^4 + \frac{1}{8} \delta^4 c_q^2 \right)
\end{align*}

(7)

We use this expansion when computing the posterior for the spherical score. Note

\begin{align*}
\quad
\left( 1 - \frac{1}{2} \delta^2 c_q + \frac{1}{8} \delta^4 c_q^2 \right) &= \frac{1}{2} \left( 2 - \delta^2 c_q + \frac{1}{4} \delta^4 c_q^2 \right) \\
&= \frac{1}{2} \left( 1 + 1 - \delta^2 c_q + \frac{1}{4} \delta^4 c_q^2 \right) \\
&= \frac{1}{2} \left( 1 + \left( \frac{1}{2} \delta^2 c_q - 1 \right)^2 \right)
\end{align*}

and so $f_q^a$ in (7) is always positive. Ignoring terms $\delta^4$ and higher gives the simpler expression

\begin{align*}
\quad
f_q^a &= c \left( 1 + \frac{1}{2} \delta^2 \left[ (\omega_q')^2 \right] / \delta^2 - c_q \right).
\end{align*}

and this approximation is suitable for deriving local procedures that are uniform in $\delta^2$.

*Derivation of the Orders of Magnitude*

Consider the regression where $X$ consists of a column of 1’s and the variable $x_i$ i.e. $X = [1 \ x]$. Define $\Delta = \sum_{j=1}^T (x_j - \bar{x})^2$, $x_2 = \sum_{j=1}^T x_j^2 / T$ and vectors $\bar{x} = x_2 \mathbf{1} - \bar{x} \mathbf{1}$, $\hat{x} = x - \bar{x} \mathbf{1}$ with $\mathbf{1}$ being a column of 1’s. Then

\begin{align*}
\quad
I - X \left( X' X \right)^{-1} X' &= I - \Delta^{-1} \left( \bar{x} \mathbf{1}' + \hat{x} \mathbf{x}' \right)
\end{align*}
In the case of the broken trend, with $b_q = (0, \ldots, 0, 1, \ldots, (T - q + 1)$, and substituting the trend notation $t = x$ (with $i$ as its mean) we get

$$c_q = b_q' b_q - t_2 \Delta^{-1} (1'b_y)^2 - \Delta^{-1} (t'b_y)^2 + 2\Delta^{-1} 1'b_q b't.$$ 

Hence using $\approx$ to denote orders of magnitude and with $\sum_{t=1}^T t^2 \approx \frac{1}{3} T^3$ and $\sum_{t=1}^T t \approx \frac{1}{2} T^2$ we see

$$c_q \approx \frac{(T-q)^3}{3} - \frac{T^2}{3} \frac{12}{T^3} \left( \frac{(T-q)^2}{2} \right)^2 - \frac{12}{T^3} \frac{1}{6} (T-q)(T-q)(2T+q) \left( \frac{(T-q)^2}{2} \right)$$

$$+ T \frac{12}{T^3} \frac{1}{6} (T-q)(T-q)(2T+q) \frac{(T-q)^2}{2}$$

$$= \frac{1}{3T^3} q^3 (T-q)^3$$

$$= \frac{1}{3} T^3 \tau^3 (1 - \tau)^3$$

with $\tau = q/T$. Let $p = \max(q, r)$, $\mu = s/T$ and noting that $b_q' t = \sum_{j=1}^{T-q} j (q + j) = \frac{1}{6} (T-q)(T-q+1)(2T+q+1) \approx \frac{1}{6} (T-q)^2 (2T+q)$ we find

$$c_{q,s} = b_q' b_q - t_2 \Delta^{-1} b_q' 11'b_y + \Delta^{-1} b_q' t_1t'b_y - \Delta^{-1} b_q' t't'b_y + \Delta^{-1} b_q' t't'b_y$$

$$\approx \frac{(T-p)^3}{3} + |s-q| \frac{(T-p)^2}{2}$$

$$- t_2 \frac{12}{T^3} \left( \frac{(T-q)^2}{2} \right) \left( \frac{(T-s)^2}{2} \right) + \frac{T}{2} \frac{12}{T^3} \frac{1}{6} (T-q)^2 (2T+q) \left( \frac{(T-s)^2}{2} \right)$$

$$- \frac{12}{T^3} \left( \frac{1}{6} (T-q)^2 (2T+q) \right) \left( \frac{1}{6} (T-r)^2 (2T+s) \right) + \frac{T}{2} \frac{12}{T^3} \frac{1}{6} (T-q)^2 (2T+s)$$

$$= \frac{1}{3} (T-p)^3 + |s-q| \frac{(T-p)^2}{2} - \frac{1}{6T^3} (T-q)^2 (T-s)^2 (T_q + T_s + 2qs + 2T^2)$$

$$= \frac{1}{6} T^3 \left( 3 |\mu - \tau| - 2 \max(\tau, \mu) + 2 \right) - (1-\tau)^2 (1-\mu)^2 (\tau + \mu + 2\tau\mu + 2)$$

\section*{References}

Andrews, D. W. K., Lee, I. and Ploberger, W. (1996). ‘Optimal Change Point Tests for Normal Linear Regression’, \textit{Journal of Econometrics}, Vol. 70, pp. 9–38.

Bai, J. (1997). ‘Estimation of a Change Point in Multiple Regressions’, \textit{Review of Economics and Statistics}, Vol. 79, pp. 551–563.

Bai, J. and Perron, P. (1998). ‘Estimating and Testing Linear Models with Multiple Structural Changes’, \textit{Econometrica}, Vol. 66, pp. 47–78.
Deng, A. and Perron, P. (2008). ‘A Non-Local ‘Perspective on the Power Properties of the CUSUM and CUSUM of Squares Tests for Structural Change’, *Journal of Econometrics*, Vol. 142, No. 1, pp. 212–240.

Elliott, G. and Muller, U. K. (2007). ‘Confidence Sets for the Date of a Single Break in Linear Time Series Regressions’, *Journal of Econometrics*, Vol. 141, pp. 1196–1218.

Ferguson, T. S. (1961). ‘On the Rejection of Outliers’, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, pp. 253–287.

Ferguson, T. S. (1967). Mathematical Statistics, New York, Academic Press.

Forchini, G. (2012). ‘Optimal Similar Tests for Structural Change for the Linear Regression Model’, *Econometric Theory*, Vol. 18, pp. 853–867.

Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). ‘Probabilistic Forecasts’, *Calibration and Sharpness*, *Journal of Royal Statistical Society B*, Vol. 69, pp. 243–268.

McCabe, B. P. M. and Harrison, M. J. (1980). ‘Testing the Constancy of Regression Relationships Over Time Using Least Squares Residuals’, *Applied Statistics*, Vol. 29, pp. 142–148.

Pitman, E. (1937). ‘The "closest" Estimates of Statistical Parameters’, *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 33, No. 2, pp. 212–222.

Ploberger, W. and Krämer, W. (1990). ‘The Local Power of the CUSUM and CUSUM of Squares Tests’, *Econometric Theory*, Vol. 6, pp. 335–347.

Ploberger, W. and Krämer, W. (1992). ‘The CUSUM Test with OLS Residuals’, *Econometrica*, Vol. 60, pp. 271–285.

Xu, K. L. (2013). ‘Powerful Tests for Structural Change in Volatility’, *Journal of Econometrics*, Vol. 173, No. 1, pp. 126–142.