

Change points and structural breaks

by Kevin Kotzé

When performing an analysis that involves time series variables, we would usually want to make use of a sample period that has a consistent data generating process, when using a traditional linear regression model. To ensure that our selected sample period satisfies this criterion we could make use of a change point or structural break test to identify a change in the underlying data generating process. Since events such as the Global Financial Crisis and the Covid-19 pandemic, there is a growing need to be able to identify the location of multiple change points within time series.

The literature on this area of work is extensive and recent advances consider the use of several different models at a level of generality that allow a host of interesting practical applications. These include models with stationary regressors and errors that can exhibit temporal dependence and heteroskedasticity, models with trending variables and possible unit roots and cointegrated models, among others.

These procedures may be used to test for common breaks across a large number of variables in either the mean or the variance of a time series or a component thereof, as well as changes in forecast accuracy that may be due to a change in the underlying data generating process. These models have made use of various estimation strategies that include ordinary least-squares, instrumental variables, Bayesian techniques, quantile regressions and methods based on the Least Absolute Shrinkage and Selection Operator (LASSO), which are usually applied within the context of a time series, panel, or factor models.

In what follows, we will only focus on specific aspects that relate to econometric applications that are based on linear relationships between variables. Our focus is on making use of retrospective (*offline*) methods that test for breaks in a given sample of data and form confidence intervals around the break dates. For recent reviews of the literature, see Perron (2006), Perron (2010), Casini and Perron (2019) and Burg and Williams (2020), while the webpage <http://changeoint.info> (<http://changeoint.info>) includes a number of important references.

1 Change point detection

A particular change point test seeks to identify the specific period of time that relates to a change in the probability distribution of a stochastic process or time series. In general, the problem concerns both detecting whether or not a change has occurred, or whether several changes might have occurred, and identifying the times of any such changes.

To establish a relatively general framework for the detection of a change point, we can assume that we have an ordered sequence of data, $y_t = \{y_1, \dots, y_T\}$. A change point is then said to arise within this dataset when there exists a time, $\tau \in \{1, \dots, T-1\}$, such that the statistical properties of $\{y_1, \dots, y_\tau\}$ and $\{y_{\tau+1}, \dots, y_T\}$ are different in some way. Extending this idea for a single change point to multiple changes, we can allow for m change points, that are associated with positions,

$\tau_{1:m} = \{\tau_1, \dots, \tau_m\}$. Each change point position is then ordered so that $\tau_i < \tau_j$ if, and only if, $i < j$. Consequently the m change points will split the data into $m + 1$ segments, where the i th segment may be summarised by a set of parameters. The parameters associated with the i th segment will be denoted $\{\theta_i, \phi_i\}$, where ϕ_i is a possible set of nuisance parameters and θ_i is the set of parameters that may describe the change. In this case, we typically want to test how many segments are needed to provide the best representation of the data generating process.

1.1 Single change point detection

Before considering the more general problem of identifying the position of multiple $\tau_{1:m}$, change points, we first consider the identification of a single change point with the aid of a likelihood based framework. The detection of a single change point can be posed as a hypothesis test, where the null hypothesis, H_0 , corresponds to no change point ($m = 0$) and the alternative hypothesis, H_1 , would suggest that there is a single change point ($m = 1$).

We now introduce the general likelihood ratio based approach to test this hypothesis. The potential for using a likelihood based approach to detect change points was first proposed by Hinkley (1970), who derives the asymptotic distribution of the likelihood ratio test statistic for a change in the mean within normally distributed observations. This approach has since been extended to consider changes in variance within normally distributed observations by Gupta and Tang (1987). The interested reader is referred to Silva and Teixeira (2008) and Eckley, Fearnhead, and Killick (2011) for a more comprehensive review.

Using the likelihood ratio method, we can construct a test statistic to decide whether a change has occurred after calculating the maximum log-likelihood under both null and alternative hypotheses. For the null hypothesis the maximum log-likelihood is $\log p(y_{1:T}|\hat{\theta})$, where $p(\cdot)$ is the probability density function associated with the distribution of the data and $\hat{\theta}$ is the maximum likelihood estimate for the parameters.

Under the alternative hypothesis, where a change point occurs at position τ_1 , with $\tau_1 \in \{1, 2, \dots, T - 1\}$, the maximum log-likelihood for a given τ_1 is

$$ML(\tau_1) = \log p(y_{1:\tau_1}|\hat{\theta}_1) + \log p(y_{\tau_1+1:T}|\hat{\theta}_2) \quad (1.1)$$

Given the discrete nature of the change point location, the maximum log-likelihood value under the alternative is simply $\max_{\tau_1} ML(\tau_1)$, where the maximum is taken over all possible change point locations. The test statistic is thus

$$\lambda = 2 \left[\max_{\tau_1} ML(\tau_1) - \log p(y_{1:T}|\hat{\theta}) \right] \quad (1.2)$$

The test involves choosing a threshold, c , such that we reject the null hypothesis if $\lambda > c$. If we reject the null hypothesis, then we could provide an estimate its position as $\hat{\tau}_1$, which relates to the position of τ_1 that maximises $ML(\tau_1)$. The appropriate value for the threshold parameter, c , is an open research question with several authors devising p values and other information criteria under different types of changes. We refer the interested reader to Guyon and Yao (1999), Chen and Gupta (2000), Lavielle (2005), Birgé and Massart (2006) for discussions and suggestions for c .

It is clear that the likelihood test statistic can be extended to multiple changes simply by summing the likelihood for each of the possible m segments. The problem then becomes one of identifying the maximum of $ML(\tau_1)$ over all possible combinations of τ_1 and m .

1.2 Multiple change point detection

Killick and Eckley (2014) suggest that the most common approach to identify multiple change points in the literature is to minimise the following expression:

$$\sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i})] + \beta f(m) \quad (1.3)$$

where \mathcal{C} is a cost function for a segment. In this case $\beta f(m)$ is a penalty to guard against over fitting which may be combined with a negative log-likelihood (i.e. a multiple change point version of the threshold c). Standard penalty functions include various forms of information criteria, such as the Schwarz, Bayesian, Akaike, or Hannan-Quinn criteria, as well as a number of additional penalty functions that have been specifically developed for this purpose. A brute force approach to solve this minimisation problem considers 2^{T-1} solutions, reducing to $\binom{T-1}{m}$ if m is known. In what follows, we consider three multiple change point algorithms that minimise equation (1.3): binary segmentation, segment neighborhoods, and pruned exact linear time (PELT).

Killick and Eckley (2014) suggest that binary segmentation is arguably the most widely used multiple change point search method and originates from the work of Edwards and Cavalli-Sforza (1965), Scott and Knott (1974) and Sen and Srivastava (1975). When using this approach, we initially apply a single change point test to all the available data, if a change point is identified the data is split into two at the change point location. The single change point procedure is repeated on the two new data sets, before and after the change. If change points are identified in either of the new data sets, they are split further. This process continues until no change points are found in any parts of the data. This procedure is an approximate minimization of equation (1.3), with $f(m) = m$, as any change point locations are conditional on change points identified previously. Binary segmentation is thus an approximate algorithm that is computationally expedient as it only considers a subset of the 2^{T-1} possible solutions. The computational complexity of the algorithm is $\mathcal{O}(T \log T)$ but this speed can come at the expense of accuracy of the resulting change points (see Killick, Fearnhead, and Eckley (2012) for details).

The segment neighborhood algorithm was proposed by Auger and Lawrence (1989) and is applied in the application of Bai and Perron (1998). The algorithm minimises the expression given in equation (1.3) with the aid of a dynamic programming technique to obtain the optimal segmentation for $m + 1$ change points reusing the information that was calculated for the m change points. This reduces the computational complexity from $\mathcal{O}(2^T)$ for a naive search to $\mathcal{O}(QT^2)$ where Q is the maximum number of change points to identify. While this algorithm is exact, the computational complexity is considerably higher than that of binary segmentation.

Like the segment neighborhood algorithm the PELT algorithm, which was proposed in Killick, Fearnhead, and Eckley (2012), also provides an exact solution (i.e. it is not an approximation). It has

been shown to be more computationally efficient, as it makes use of both dynamic programming and pruning, which can result in an $\mathcal{O}(T)$ search algorithm subject to certain assumptions being satisfied. Most of these conditions are not particularly onerous.

1.3 Changes in mean and variance

Early work on change point problems focused on identifying changes in mean and includes the work of Page (1954) and Hinkley (1970) who created the likelihood ratio and cumulative sum (CUSUM) test statistics, respectively. Although tests for changes in variance have received less attention much of the work in this area builds on the work of Hinkley (1970). This extension is discussed in Hsu (1979), Horvath (1993) and Chen and Gupta (1997), while Killick et al. (2010) note that it is usually relatively challenging to detect subtle changes in variability. One is also able to test for a combined change in both the mean and variance, when the data takes on an assumed distribution.

To provide an example of the output from these tests, we can make use of simulated data, where in the initial case we are going to draw 360 observations from a random normal distribution with a constant variance. The first 100 observations have a mean value of zero, the second 50 observations have a mean of 1.5, the third 90 observations have a mean of zero and the last 120 have a mean value of -0.8. This data is displayed in Figure 1.

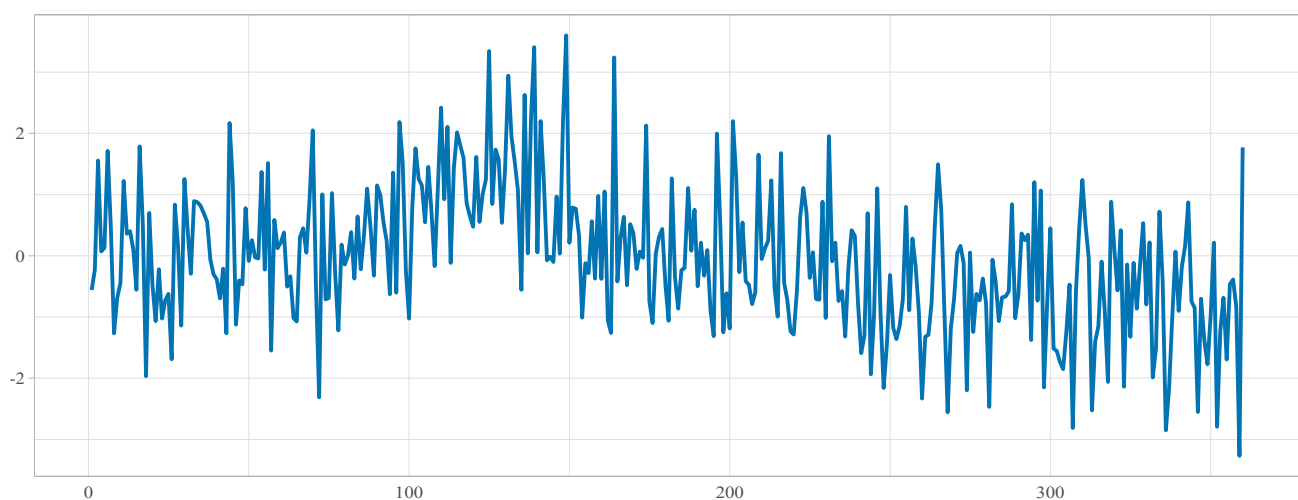
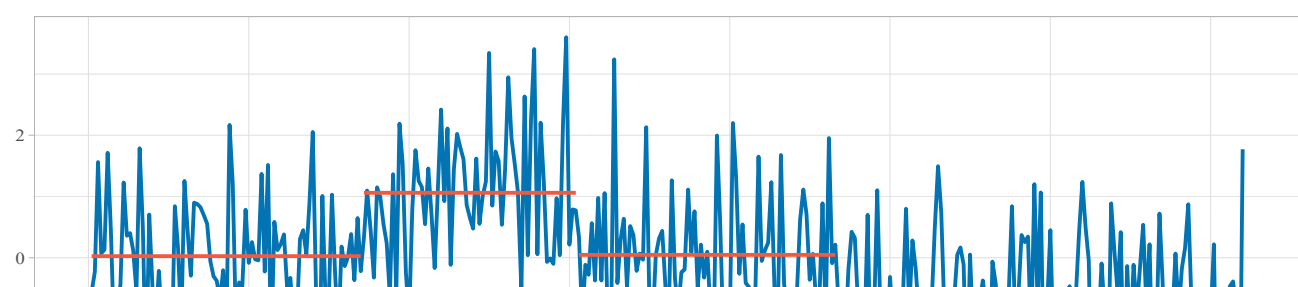
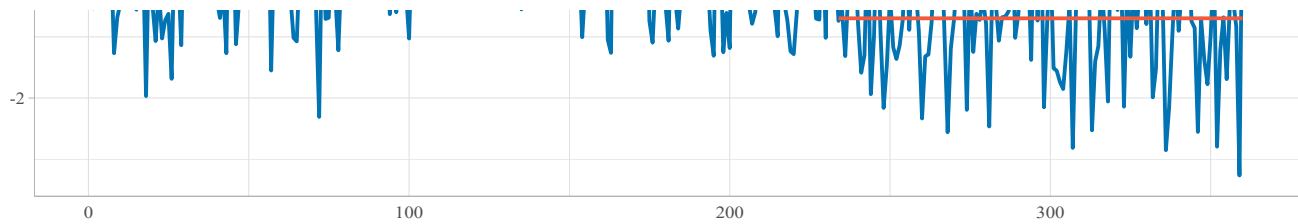


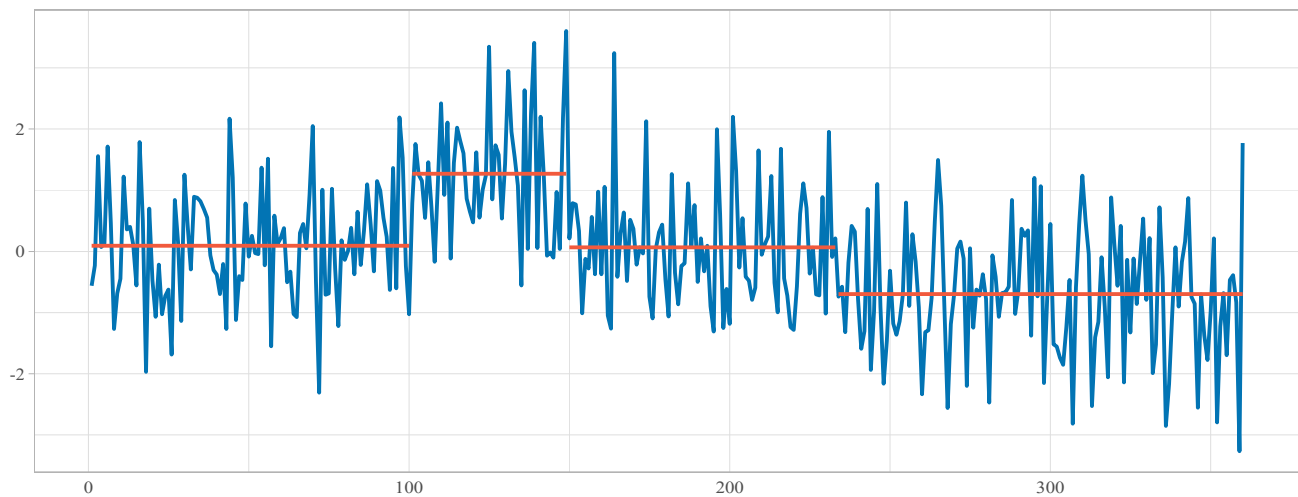
Figure 1: **Simulated time series with different means**

Using the binary segmentation method, it suggests that by segmenting the data after observations 85, 152, and 233, we would be able to improve our description of the data, as suggested by the Bayesian information criteria. Figure 2 displays the result for the change in the mean, when using this method.

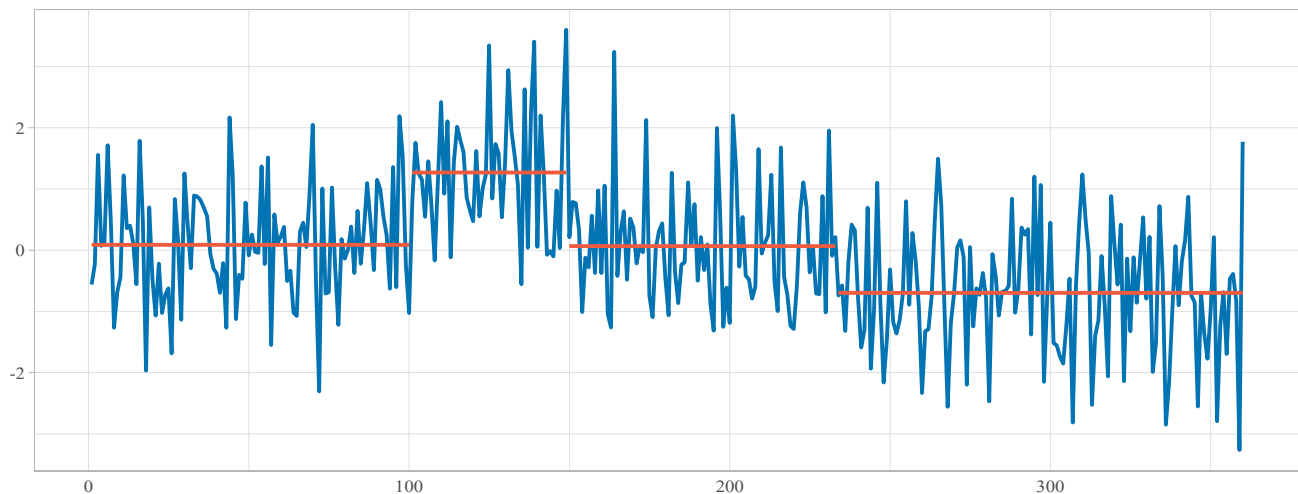


Figure 2: **Binary segmentation method**

While the binary segmentation method provides the most expedient result, we can provide a more accurate result with the segment neighborhood algorithm, which suggest that the breaks arise at observations 100, 149 and 233. This result is displayed in Figure 3 displays the result for the change in the mean, when using this method.

Figure 3: **Segment neighborhood method**

And the same result is produced when using the more expedient pruned optimisation algorithm. This result is displayed in Figure 4.

Figure 4: **PELT method**

To the subsequent draw of 360 observations from a random normal distribution we make use of the same mean but different variances. The first 100 observations have a variance of 1, the second 50 observations have a variance of 2, the third 90 observations have a variance of 1 and the last 120 have a variance of 0.5. After making use of the PELT algorithm we are able to identify change points in variance at observations 104, 150 and 236 to provide the results that are displayed in Figure 5.

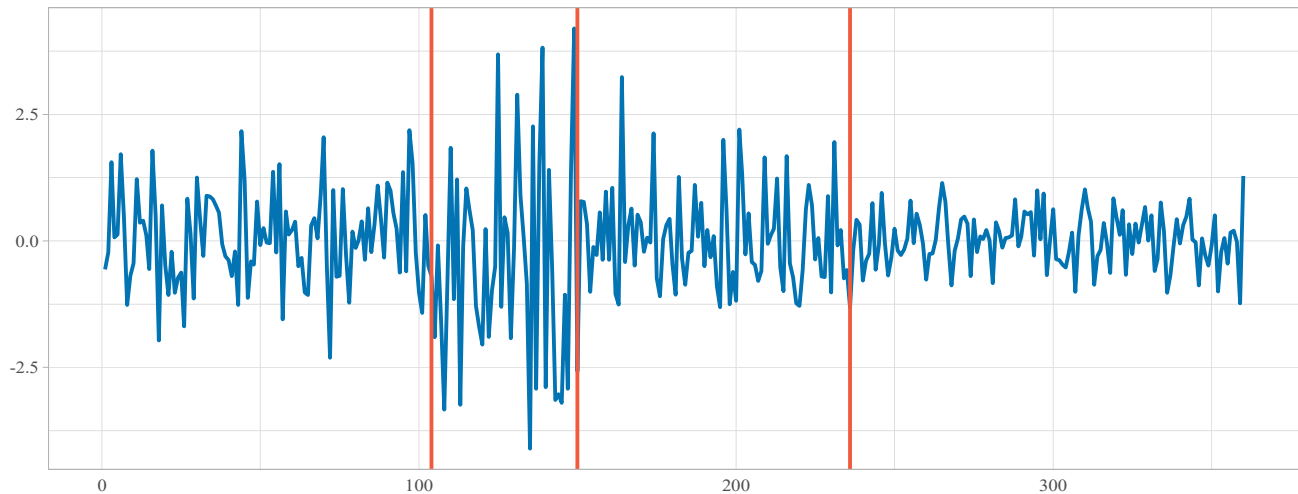


Figure 5: **Identifying a change in variance**

To identify a change in both the mean and variance we are going to assume that the above changes in the mean and variance arise simultaneously and that the data is normally distributed. After performing this test we are able to identifying change points at observation 107, 150 and 239 to produce results that are displayed in Figure 6.

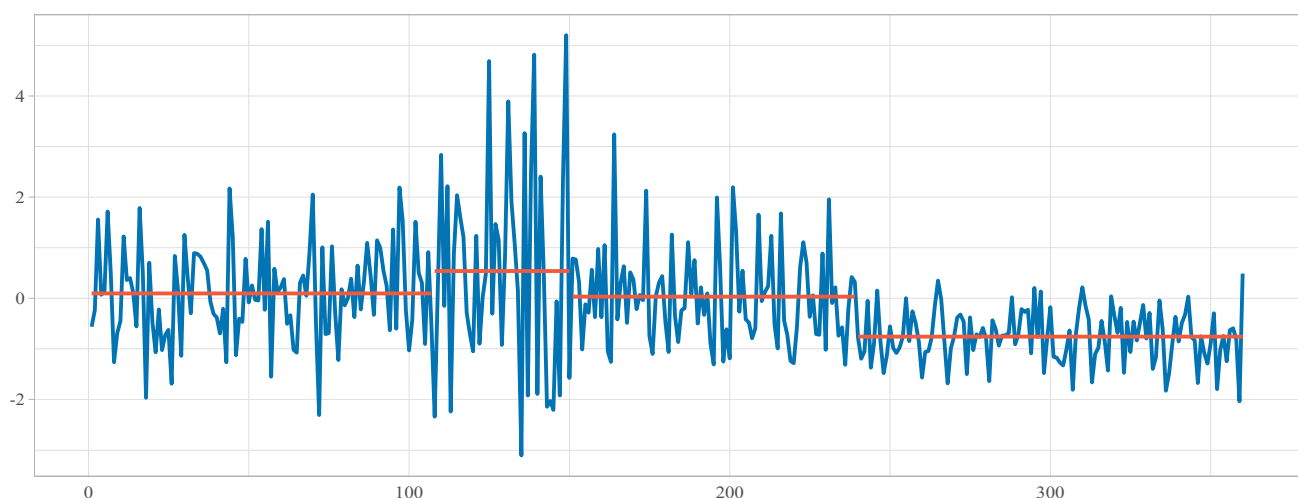


Figure 6: **Identifying a change in the mean and variance**

2 Structural break tests

The term structural break is in many respects synonymous with change point, but would refer to specific cases where there is a change in the regression coefficients. Therefore, a structural break test

is subject to the particular specification of model that is under consideration. In the case of a linear regression model:

$$y_t = x_t^\top \beta_j + \varepsilon_t \quad j = \{1, \dots, m+1\}$$

the hypothesis that the regression coefficients are constant may be constructed as follows:

$$H_0 : \beta_j = \beta_0$$

against the alternative that at least one coefficient varies over time. In the case where x_t takes the form of a constant, then the above test would be equivalent to testing for a change point in the mean.

To test such a hypothesis we could also make use of information that is contained in the residual, ε_t . Such a structural break test could make use of F -statistics or generalised fluctuation tests to compare the null hypothesis against the alternative. The F -statistics consider whether or not a change occurred at time τ , where the residuals for a subsample $\hat{\varepsilon}(\tau)$ are compared to the residuals for the unsegmented dataset, $\hat{\varepsilon}$, where

$$F_i = \frac{\hat{u}^\top \hat{u} - \hat{u}(\tau)^\top \hat{u}(\tau)}{\hat{u}(\tau)^\top \hat{u}(\tau) / (T - 2k)}$$

When the date of the potential structural break is unknown, one would test all possible positions for τ , where the null hypothesis is rejected if the supremum is too large. Hansen (1997) provides critical values that are used for the approximate asymptotic p values for this test.

The generalized fluctuation test framework seeks to identify departures from constancy in a graphical way, where after fitting a particular regression model to the data, before considering the behaviour of fluctuations in either the residuals or the parameter estimates. For example, variants of the CUSUM test that was introduced by Brown, Durbin, and Evans (1975), considers the cumulated sums of the residuals from a particular model. If they represent white noise then it would be expected that they should be centered on zero. However, where these residuals display a significant departure from zero, then this behaviour may suggest that there is a structural break in the data.

2.1 Chow's breakpoint test

The Chow (1960) breakpoint test seeks to fit the same regression model to separate sub-samples of the data, to see whether there are significant differences in the parameter estimates. Formally, this test investigates whether the null hypothesis of “no structural change” holds after constructing a F -test statistic for the parameters in the two models. A significant difference indicates a structural change in the relationship.

To gain the intuition behind this statistic, consider the simple linear regression model,

$$y_t = x_t^\top \beta_j + \varepsilon_t$$

where ε_t is a serially uncorrelated error term. In this case, we allow for the possibility that β_j may be time varying, in that it may take on two possible values. To test whether or not the coefficient estimate changes at date τ , we could consider the values for β_j , where

$$\beta_j = \begin{cases} \beta & t \leq \tau \\ \beta + \delta & t > \tau \end{cases}$$

If the break date is known, then the problem of testing the null hypothesis of no break (that is, $\delta = 0$) against the alternative of a nonzero break ($\delta \neq 0$) is equivalent to testing the hypothesis that the coefficient δ is zero in the augmented regression

$$y_t = x_t^\top \beta_t + \delta Z_t(\tau) + \varepsilon_t$$

where $Z_t(\tau) = x_t$ if $t > \tau$ and $Z_t(\tau) = 0$ otherwise.

This test can be computed using a conventional t -statistic when estimating the regression with ordinary least squares. The hypothesis of no break is rejected at the 5% significance level if the absolute value of this t -statistic is greater than 1.96. Or alternatively, we could test for a break in all the model parameters at this point in time with the aid of an F -statistics, as described above.

The major drawback of this procedure is that the change point must be known *a priori* as it is required to split the sample into two sub-samples. In addition, to perform this test you would need to ensure that each sub-sample has at least as many observations as the number of estimated parameters.

2.2 Quandt likelihood ratio test

The Quandt (1960) likelihood ratio (QLR) test is a natural extension of the Chow test where a F -test statistic is calculated for all potential breakpoints within a particular interval. This interval is usually dependent upon the degrees of freedom that are required for the estimation of the regression model. The largest test statistic across the grid of all potential break points is then identified as the QLR statistic, as it indicates the most likely break point. One would then reject the null hypothesis of no structural change if the absolute value of this test statistic is relatively large. Andrews (1993) and Andrews and Ploberger (1994) developed an applicable distribution for this test-statistic, which is used in for the calculation of asymptotic p -values. Alternative critical values are discussed in Hansen (1997) and Stock and Watson (2010). It is usually performed as a $\sup F$ -test although other variants exist. This statistic displays good power against the alternative of a breakpoint.

2.3 Testing for multiple structural breaks

Bai and Perron (1998) and Bai and Perron (2003) extend this approach to test for multiple structural breaks. They make use of F -tests for 0 vs. τ_1 breaks and τ_1 vs. τ_2 breaks, etc. Hence, where it is assumed that there are m possible breakpoints they consider the least squares estimates for the different possible β_j coefficients to identify the group of coefficients that would minimise the resulting residual sum of squares.

To implement this procedure Bai and Perron (2003) calculate the residual sum of squares from a regression that includes a single breakpoint, for which the date of which is not observed. The value for this statistic is then compared to the residual sum of squares for the regression that does not include a breakpoint. Thereafter, a second breakpoint is included and the residual sum of squares is compared to that of a single breakpoint. Therefore, the problem for identifying all the dates for the

structural changes is to find the m breakpoints that minimise the residuals sum of squares over all m partitions. Information criteria are often used for model selection, which would identify the selection of m breakpoints in this case. Bai and Perron (2003) suggest that the AIC usually overestimate the number of breaks and as such the BIC is usually preferred. To perform this calculation in a relatively expedient manner they make use a dynamic programming procedure.

2.4 CUSUM test

The CUSUM test is based on the cumulative sum of the recursive residuals that utilises a generalised fluctuation test framework. To investigate for a structural break, one would plot the cumulative sum together with the 5% critical boundaries. The test finds parameter instability if the cumulative sum breaks either of the two boundaries. It was originally proposed by Brown, Durbin, and Evans (1975) to test the null hypothesis of parameter stability and is particularly useful when the process is initially relatively stable, and then goes through a relatively turbulent period.

2.5 Examples of structural breaks

To provide an example of the output from the above structural break tests, we can make use of simulated data once again. However, in this case we are going to assume that the data displays a certain degree of persistence, where the first 100 observations represent an autoregressive process with a single coefficient of 0.9, the second 100 observations represent a moving average process with a coefficient of 0.1, and the last 100 observations represent and ARMA(1,1) with coefficients of 0.5 and 0.3. This time series variables is displayed in Figure 7.

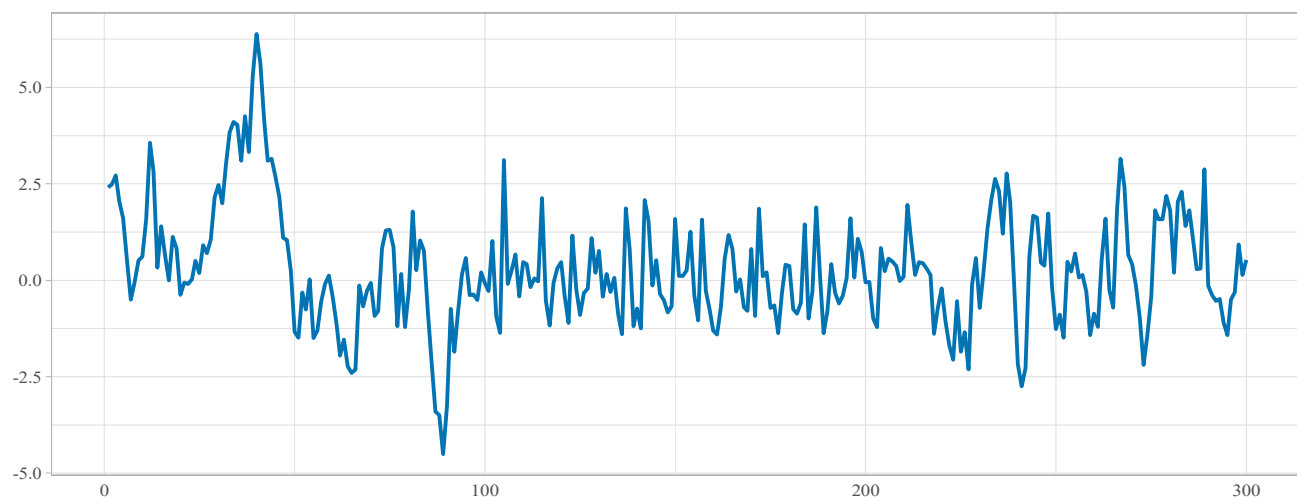
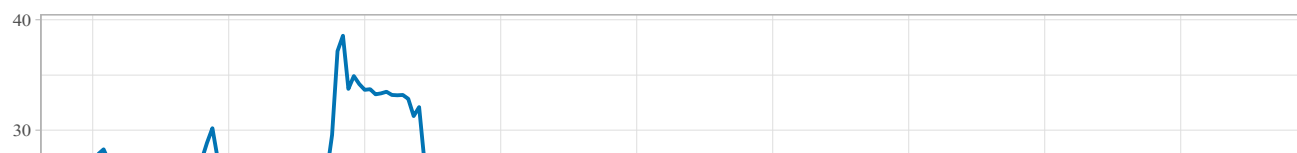
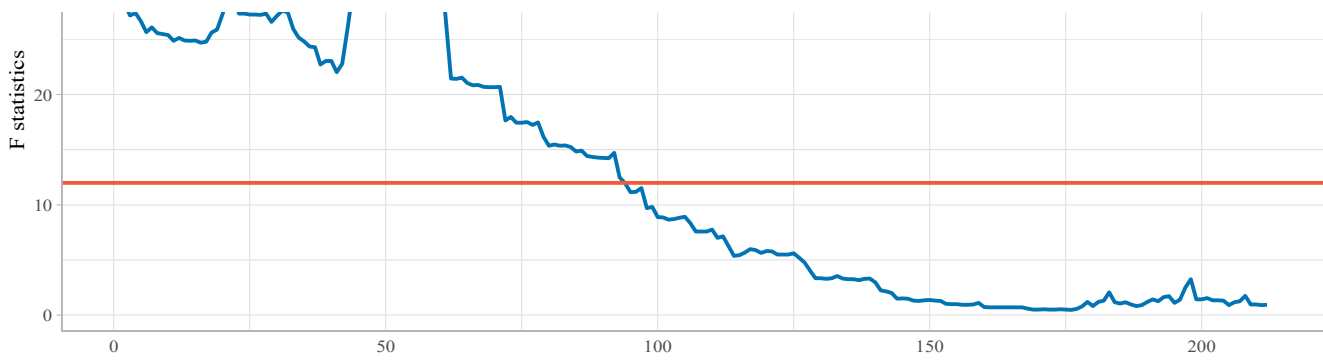


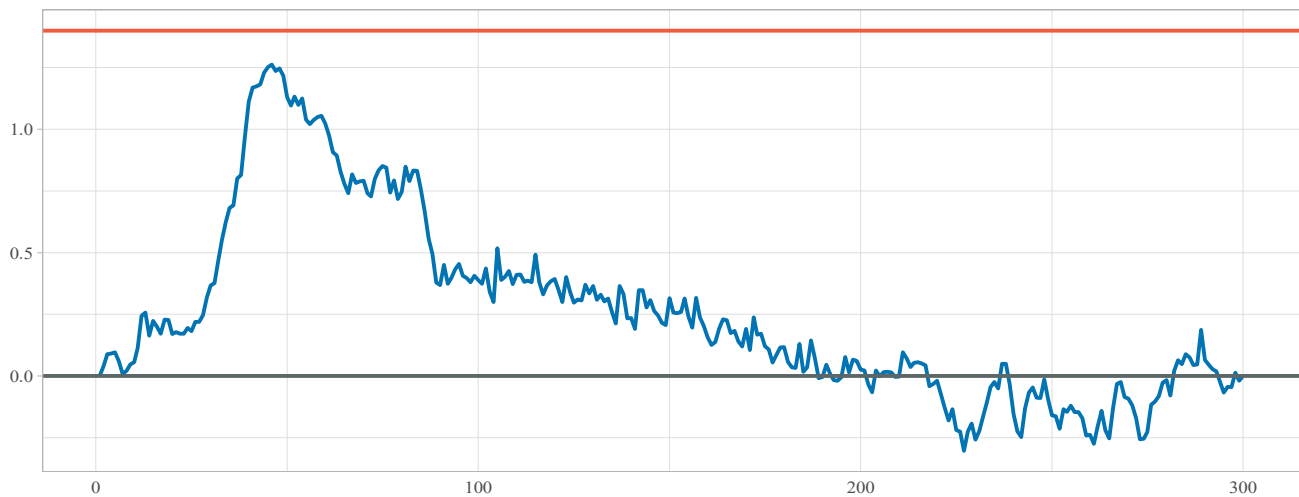
Figure 7: **Simulated time series with varying persistence**

When using the QLR method we note from the F -statistics that are displayed in Figure 8 suggest that the most likely structural break arises at observation 89.

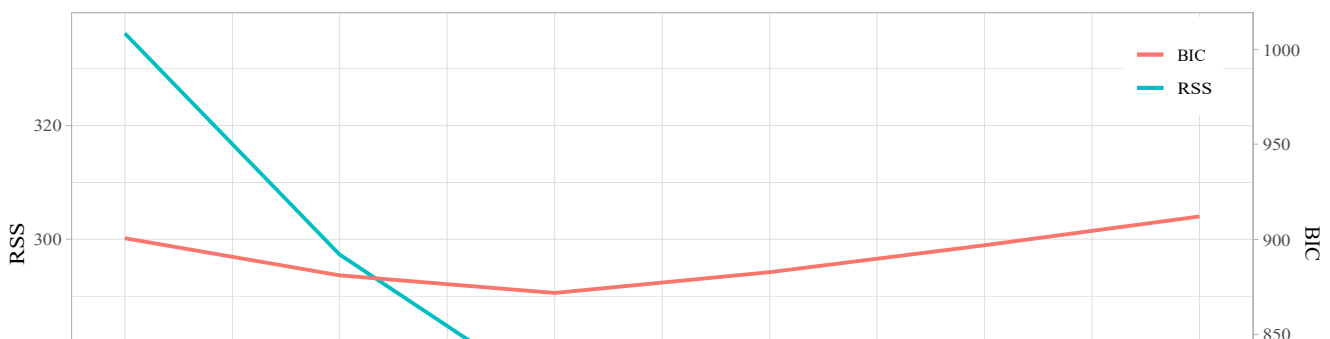


Figure 8: **Quandt likelihood ratio test**

When using the CUSUM test, there appear to be no breaks in the data that breach the critical values, which would suggest that it is not able to identify the structural breaks that are present in the data. These results are displayed in Figure 9.

Figure 9: **CUSUM test**

Then lastly, the use of the Bai and Perron (2003) method to identify multiple structural breaks, suggest that while the Bayesian information criteria initially declines after including two structural breaks, it the starts to increase with the addition of further potential structural breaks. These two structural break arise at observations 89 and 204. This result is displayed in Figure 10 where we plot measures for the RSS and BIC for the different number of structural breaks, which are displayed on the horizontal axis.



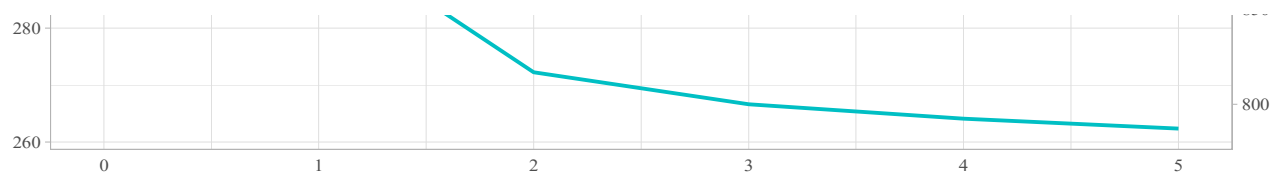


Figure 10: **Bai and Perron (2003) structural break test**

3 Conclusion

Time series models consider the relationship between variables that are observed over a particular period of time. Most of these models assume that the relationship between the variables remains constant over the entire period. Change point tests and structural break models seek to identify permanent changes that may arise in the first two moments of the data or the parameters of the models.

Several widely used economic and financial indicators have a number of potential structural breaks, particularly over recent periods of time. Failing to recognise this aspect can lead to invalid conclusions about the relevant features of the data and the forecasts from such models would usually be relatively inaccurate.

Where the first two moments of the data may potentially be influenced by a change point, one could perform one of the many tests that have been described above. In addition, when using a linear model on data that may be subject to a potential structural break, one could make use of a structural break test, where the QLR statistic or a CUSUM test may be used to identify a single break. Alternatively, to test for multiple structural breaks, one may employ the Bai and Perron (1998) or Bai and Perron (2003) methodologies.

4 References

- Andrews, D. W. K. 1993. “Tests for Parameter Instability and Structural Change with Unknown Change Point.” *Econometrica* 61(4): 821–56.
- Andrews, D. W. K., and W. Ploberger. 1994. “Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative.” *Econometrica* 62: 1383–1414.
- Auger, I. E., and C. E. Lawrence. 1989. “Algorithms for the Optimal Identification of Segment Neighborhoods.” *Bulletin of Mathematical Biology* 51 (1): 39–54.
- Bai, Jushan, and Pierre Perron. 1998. “Estimating and Testing Linear Models with Multiple Structural Changes.” *Econometrica* 66 (1): 47–78.
- . 2003. “Computation and Analysis of Multiple Structural Change Models.” *Journal of Applied Econometrics* 18 (1): 1–22.
- Birgé, Lucien, and Pascal Massart. 2006. “Minimal Penalties for Gaussian Model Selection.” *Probability Theory and Related Fields* 138 (1): 33–73.
- Brown, R. L., J. Durbin, and J. M. Evans. 1975. “Techniques for Testing the Constancy of Regression Relationships over Time.” *Journal of the Royal Statistical Society* 37: 149–63.

- Burg, Gerrit J. J. van den, and Christopher K. I. Williams. 2020. "An Evaluation of Change Point Detection Algorithms," March. <http://arxiv.org/abs/2003.06222v2> (<http://arxiv.org/abs/2003.06222v2>).
- Casini, Alessandro, and Pierre Perron. 2019. "Structural Breaks in Time Series." Oxford University Press.
- Chen, J., and A. K. Gupta. 1997. "Testing and Locating Variance Changepoints with Application to Stock Prices." *Journal of the American Statistical Association* 92 (438): 739–47.
- . 2000. *Parametric Statistical Change Point Analysis*. Birkhauser.
- Chow, G. C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28: 591–605.
- Eckley, I. A., P. Fearnhead, and R. Killick. 2011. "Analysis of Changepoint Models." In *Bayesian Time Series Models*. Cambridge University Press.
- Edwards, A. W. F., and L. L. Cavalli-Sforza. 1965. "A Method for Cluster Analysis." *Biometrics* 21 (2): 362–75.
- Gupta, A. K., and J. Tang. 1987. "On Testing Homogeneity of Variances for Gaussian Models." *Journal of Statistical Computation and Simulation* 27 (2): 155–73.
- Guyon, Xavier, and Jian-feng Yao. 1999. "On the Underfitting and Overfitting Sets of Models Chosen by Order Selection Criteria." *Journal of Multivariate Analysis* 70 (2): 221–49.
- Hansen, B.E. 1997. "Approximating Asymptotic p Values for Structural-Change Tests." *Journal of Business and Economic Statistics* 15 (1): 60–67.
- Hinkley, D. V. 1970. "Inference About the Change-Point in a Sequence of Random Variables." *Biometrika* 57 (1): 1–17.
- Horvath, L. 1993. "The Maximum Likelihood Method of Testing Changes in the Parameters of Normal Observations." *Annals of Statistics* 21 (2): 671–80.
- Hsu, D. A. 1979. "Detecting Shifts of Parameter in Gamma Sequences with Applications to Stock Price and Air Traffic Flow Analysis." *Journal of the American Statistical Association* 74 (365): 31–40.
- Killick, Rebecca, and Idris A. Eckley. 2014. "Changepoint: AnRPackage for Changepoint Analysis." *Journal of Statistical Software* 58 (3).
- Killick, R., I. A. Eckley, P. Jonathan, and K. Ewans. 2010. "Detection of Changes in the Characteristics of Oceanographic Time-Series Using Statistical Change Point Analysis." *Ocean Engineering* 37 (13): 1120–6.
- Killick, R., P. Fearnhead, and I. A. Eckley. 2012. "Optimal Detection of Changepoints with a Linear Computational Cost." *Journal of the American Statistical Association* 107 (500): 1590–8.
- Lavielle, M. 2005. "Using Penalized Contrasts for the Change-Point Problem." *Signal Processing* 85: 1501–10.

- Page, E. S. 1954. "Continuous Inspection Schemes." *Biometrika* 41 (1): 100–115.
- Perron, Pierre. 2006. "Palgrave Handbook of Econometrics, Volume 1." In, 278–352. Pagrave MacMillan.
- . 2010. "Macroeconometrics and Time Series Analysis." In, edited by S.N. Durlauf and L.E. Blume. The New Palgrave Economics Collection. London: Palgrave Macmillan.
- Quandt, R.E. 1960. "Tests of the Hypothesis That a Linear Regression Obeys Two Separate Regimes." *Journal of the American Statistical Association* 55: 324–30.
- Scott, A. J., and M. Knott. 1974. "A Cluster Analysis Method for Grouping Means in the Analysis of Variance." *Biometrics* 30 (3): 507–12.
- Sen, Ashish, and S. Srivastava. 1975. "On Tests for Detecting Change in Mean When Variance Is Unknown." *Annals of the Institute of Statistical Mathematics* 27 (1): 479–86.
- Silva, Ester G., and Aurora A.C. Teixeira. 2008. "Surveying Structural Change: Seminal Contributions and a Bibliometric Account." *Structural Change and Economic Dynamics* 19 (4): 273–300.
- Stock, James H., and Mark W. Watson. 2010. *Introduction to Econometrics*. 3rd ed. New York: Addison-Wesley.