Comments on “Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model”

XIAOLAN L. WANG

Climate Research Branch, Meteorological Service of Canada, Downsview, Ontario, Canada

28 January 2003 and 30 April 2003

Lund and Reeves (2002) revisited the two-phase linear regression test for changepoint detection at an undocumented time. The two-phase regression model is of the following form:

\[
X_t = \begin{cases} 
\mu_1 + \alpha_1 t + \epsilon, & 1 \leq t \leq c \\
\mu_2 + \alpha_2 t + \epsilon, & c < t \leq n,
\end{cases}
\]

where \( \{\epsilon_t\} \) is the zero-mean independent random error with a constant variance \( \sigma^2 \). This allows for both step- and trend-type changepoints. The time \( c \) is called a changepoint if \( \mu_1 \neq \mu_2 \) (step type) and/or \( \alpha_1 \neq \alpha_2 \) (trend type).

The \( F \) statistic for a changepoint at time \( c \in \{2, \ldots, n-1\} \) is

\[
F_c = \frac{(SSE_{\text{Red}} - SSE_{\text{Full}})/2}{SSE_{\text{Full}}/(n - 4)},
\]

where \( SSE_{\text{Full}} \) is the “full model” sum of squared errors:

\[
SSE_{\text{Full}} = \sum_{i=1}^{c} (X_i - \hat{\mu}_1 - \hat{\alpha}_1 t)^2 + \sum_{t=c+1}^{n} (X_i - \hat{\mu}_2 - \hat{\alpha}_2 t)^2
\]

and \( SSE_{\text{Red}} \) is the “reduced model” sum of squared errors:

\[
SSE_{\text{Red}} = \sum_{i=1}^{n} (X_i - \hat{\mu}_{\text{Red}} - \hat{\alpha}_{\text{Red}} t)^2,
\]

where \( \hat{\mu}_{\text{Red}} \) and \( \hat{\alpha}_{\text{Red}} \) are estimated under the constraints \( \mu_1 = \mu_2 = \mu_{\text{Red}} \) and \( \alpha_1 = \alpha_2 = \alpha_{\text{Red}} \). Under the null hypothesis of no changepoint and assuming Gaussian errors \( \epsilon_t \), \( F_c \) has an \( F \) distribution of 2 numerator degrees of freedom and \( (n - 4) \) denominator degrees of freedom, denoted as \( F_{(2, n-4)} \). The existence of an undocumented changepoint is concluded when

\[
F_{\text{max}} = \max_{1 \leq c \leq n} F_c
\]

is too large to be attributed to chance variation. The most prominent changepoint is estimated as the argument(s) \( c \) that maximizes \( F_c \). Lund and Reeves (2002) pointed out errors that had propagated in the related literature and presented the true percentiles of the \( F_{\text{max}} \) distribution under the null hypothesis of no changepoint, which is very useful for detecting step- and trend-type changepoints.

However, as mentioned in Lund and Reeves (2002), changes in trend slopes could be rooted in true climate change. Also, a trend-type changepoint could well be just a point between two phases of long quasi-periodic variation (e.g., multidecadal fluctuations). Thus, extra caution should be exercised when using the above changepoint detection technique, especially when a trend-type changepoint is involved. In particular, Eq. (4.2) in Lund and Reeves (2002), which shows how to adjust the time series for the detected changepoint, should be used with caution. If the ultimate goal of the changepoint detection is to form homogeneous climate data series by correcting biases, a trend-type change in climate data series should only be adjusted if there is sufficient evidence showing that it is related to a change at the observing station, such as a change in the exposure or location of the station, or in its instrumentation or observing procedures.

In addition, a simpler situation is most often encountered when correcting biases in climate observations but was not considered explicitly in Lund and Reeves (2002). That is a case when \( \alpha_1 = \alpha_2 = \alpha \); that is, a two-phase regression model with a common trend \( \alpha \):

\[
X_t = \begin{cases} 
\mu_1 + \alpha t + \epsilon, & 1 \leq t \leq c \\
\mu_2 + \alpha t + \epsilon, & c < t \leq n.
\end{cases}
\]
In this case, the $F$ statistic for the changepoint $c \in \{2, \ldots, n-1\}$ is

$$F_c = \frac{(\text{SSE}_{\text{Red}} - \text{SSE}_{\text{Full}})/1}{\text{SSE}_{\text{Full}}/(n-3)},$$

which is distributed as $F_{(1,n-3)}$ under the null hypothesis of no changepoint and assuming Gaussian errors $\epsilon_i$. Here,

$$\text{SSE}_{\text{Full}} = \sum_{i=1}^{n} (X_i - \hat{\mu}_1 - \hat{\alpha}t)^2 + \sum_{i=1}^{n} (X_i - \hat{\mu}_2 - \hat{\alpha}t)^2 \quad \text{and} \quad (8)$$

$$\text{SSE}_{\text{Red}} = \sum_{i=1}^{n} (X_i - \hat{\mu}_{\text{Red}} - \hat{\alpha}_{\text{Red}}t)^2, \quad (9)$$

where $\hat{\mu}_{\text{Red}}$ and $\hat{\alpha}_{\text{Red}}$ are estimated under the constraint $\mu_1 = \mu_2 = \mu_{\text{Red}}$ (i.e., $X_i = \mu_{\text{Red}} + \alpha_{\text{Red}}t + \epsilon_i$ for $1 \leq t \leq n$).

Under the null hypothesis the $F_{\max}$ percentiles of model (6) are given in Table 1, which was obtained by simulating 100 000 $F_{\max}$ values for each series length $n$. The simulation procedure is similar to the one described in Lund and Reeves (2002).

One of the advantages of using model (6) is that it has one less parameter to estimate, and hence it has smaller sampling variability (the variance of the estimates of $\alpha$ is $\text{var}(\hat{\alpha}) = \sigma^2\sum_{I}[(1 - (n + 1)/2]^2$ in this case) and a higher power of detecting step-type changepoints (i.e., more likely to detect a changepoint when it really exists). To illustrate this, for each series length $n$, we simulated 200 time series of the form $X_i = \mu + \alpha t + \epsilon_i$ with arbitrary parameters $\mu$ and $\alpha$ and $\sigma$ (note that under the null hypothesis the $F_{\max}$ statistic does not depend on the particular values of $\mu$ and $\alpha$ and $\sigma$). Then, we imposed a step $\Delta X = \sigma j/100$ at an arbitrary point $c$ ($1 < c < n$) in the $j$th simulated time series ($j = 1, 2, \ldots, 200$). In other words, the steps imposed in the 200 simulated time series range from $0.01\sigma$ to $2\sigma$.

Then, both models (6) and (1) were applied to each of the 200 time series to detect the imposed step-type changepoint. The results clearly show that model (6) has a somewhat higher power of detection than model (1), especially when the time series is short (cf. Table 2) or the “step size to noise” ratio (i.e., $\Delta X/\sigma_c$) is low (not shown). Note that each of our simulations was drawn from a model with a constant slope, in which case model (1) is not suitable and hence it is not surprising that its power of detection is lower than that of model (6).

In particular, model (6) can better identify a negative step (i.e., $\mu_1 < \mu_2$) that occurs in time series of a decreasing trend, or a positive step (i.e., $\mu_1 > \mu_2$) that occurs in time series of an increasing trend. Such steps could be “masked” and/or misrepresented as trend-type changes when using model (1). These situations were also tested in the power of detection test simulations described in the last paragraph, with results shown in Table 2. Here is a “real world” example: an observer change (from a part-time observer to a full-time observer) at Gander International Airport (Canada) on 1 April 1997 introduced a step-type changepoint between March and April of 1997 in the monthly cloud-cover time series: the full-time observer observes much more frequently in each hour and hence reports more cloudy conditions than the part-time observer does (cf. Fig. 1; the latter had other duties in addition to making observations). Both models (6) and (1) were applied to the time series of monthly counts of occurrences of 1/10 and 9/10 cloud-covered sky conditions (i.e., 1/10 and 9/10 of the sky covered by clouds). The results are shown in Fig. 1: When the step size to noise ratio is
high, such as the case of 9/10 cloud-cover (Fig. 1, bottom), both models accurately identified the changepoint, although the sizes of the step and trend estimated by the two models are different. However, when the step size to noise ratio is lower, such as the case of 1/10 cloud cover (Fig. 1, top), model (6) clearly outperforms model (1): while the changepoint identified by model (6) is identical to the actual changepoint, the one identified by model (1) is 39 time intervals (months in this case) earlier and involves a notable trend-type change, which is far from the reality.

Having detected a changepoint using model (6), one can further test whether the slopes are equal before and after the changepoint time. A standard normal $z$-score hypothesis test with $z = (\hat{a}_1 - \hat{a}_2)/[\text{var}(\hat{a}_1) + \text{var}(\hat{a}_2)]^{1/2}$ would suffice. But again, a trend-type change should only be adjusted if there is sufficient evidence showing that it is not rooted in true climate change/variability.

Finally, note that both models (6) and (1) can easily accommodate reference series (see Lund and Reeves 2002 for more details).

Acknowledgments. The author thanks Dr. Francis W. Zwiers for his helpful comments on an earlier version of this manuscript. The reviewer Dr. Robert Lund is also acknowledged. The research was partly funded by the Action Plan 2000 science program of the Canadian government.

REFERENCES