Estimating the slope and standard error of a long-term linear trend fitted to adjusted annual temperatures

Brett Mullan
National Institute of Water and Atmospheric Research
New Zealand

Stephen Stuart
National Institute of Water and Atmospheric Research
New Zealand

Peter Thomson
Statistics Research Associates Ltd
New Zealand

Outline

1. Introduction
2. Model framework
3. Global adjustment
4. Local adjustment
5. Fitting a linear trend to adjusted data
6. Conclusions

1. Introduction

An accurate understanding of the long-term evolution of temperature is key to understanding the impact of global warming.

However temperatures at a given site

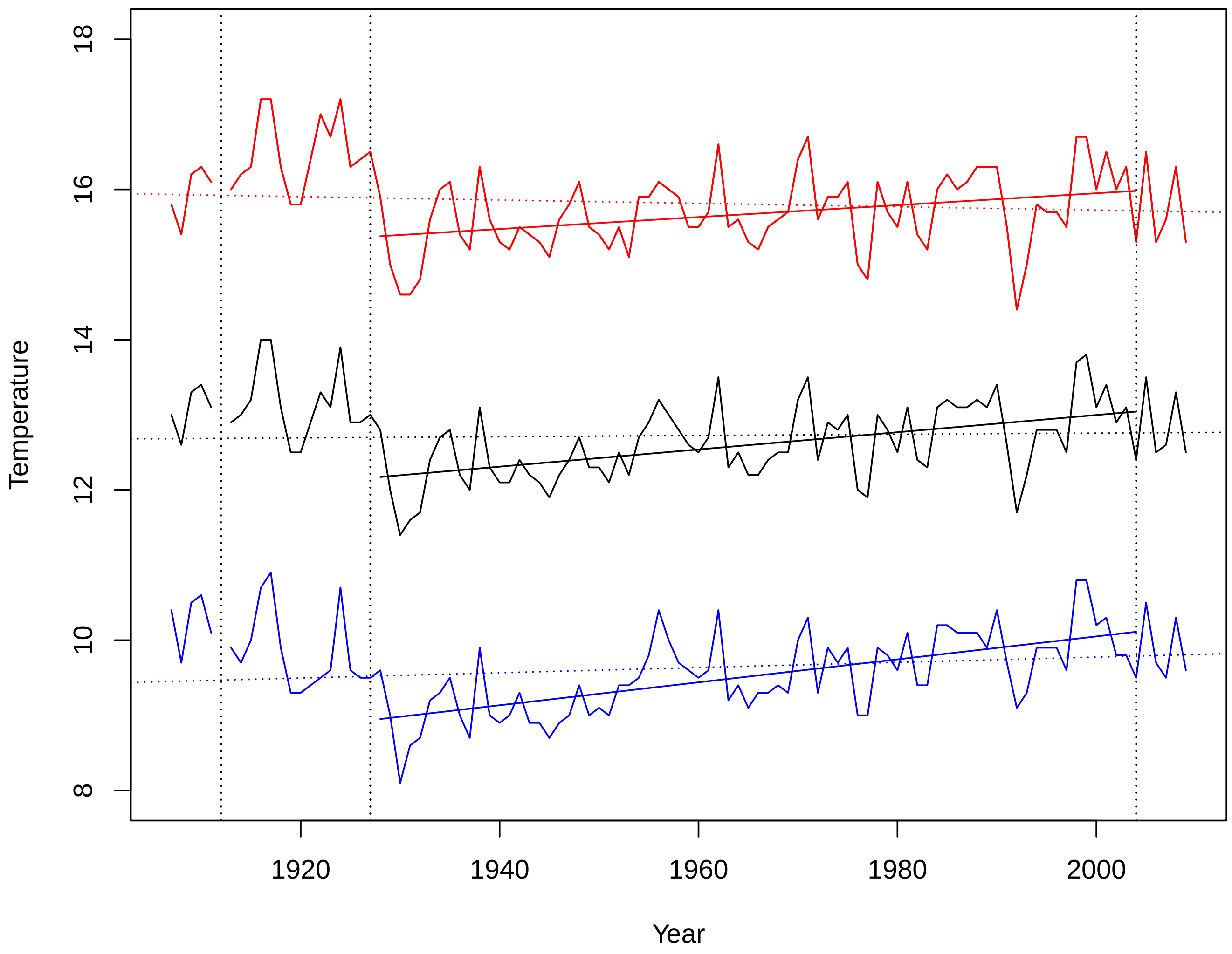
- rarely come as one homogeneous time series;
- typically comprise a collection of separate time series segments

due to re-siting recorders, changes in instrumentation etc.

Temperature segments are typically sewn together by adjusting for level shifts and other factors (homogenisation).

Adjustments are often ignored with linear trends fitted to homogenised temperatures as if they were one long homogeneous time series.

What is the impact of adjustment on estimated slopes and standard errors?



Annual average daily maximum, **mean** and minimum temperatures at Wellington, New Zealand. The OLS linear trends for the combined temperatures (dotted) and the central segment (solid) are superimposed.

2. Model framework

Focus on annual average temperatures

$$Y_k(t) = \text{average temperature for year } t \text{ at site } k$$

over a period of T years and a region of $K+1$ sites ($k = 0, \ldots, K$).

For Site 0 suppose

$$Y_0(t) = \mu_0 + T(t) + \sum_{j=1}^{p} \delta_j d(t - t_j) + \epsilon_0(t)$$

where

- μ_0 is a temperature offset;
- $T(t)$ is the common regional temperature signal;
- the δ_j are level shifts with $d(t) = 1$ $(t \leq 0)$ and 0 otherwise;
- the changepoints t_j are known;
- the measurement errors $\epsilon_0(t)$ are white noise and uncorrelated with $T(t)$.

Further assume that $T(t)$ is trend-stationary with

$$T(t) = \alpha + \beta t + X(t)$$

where $X(t)$ is stationary with mean zero. Now

$$Y_0(t) = \mu_0 + \beta t + \sum_{j=1}^{p} \delta_j d(t - t_j) + e_0(t)$$

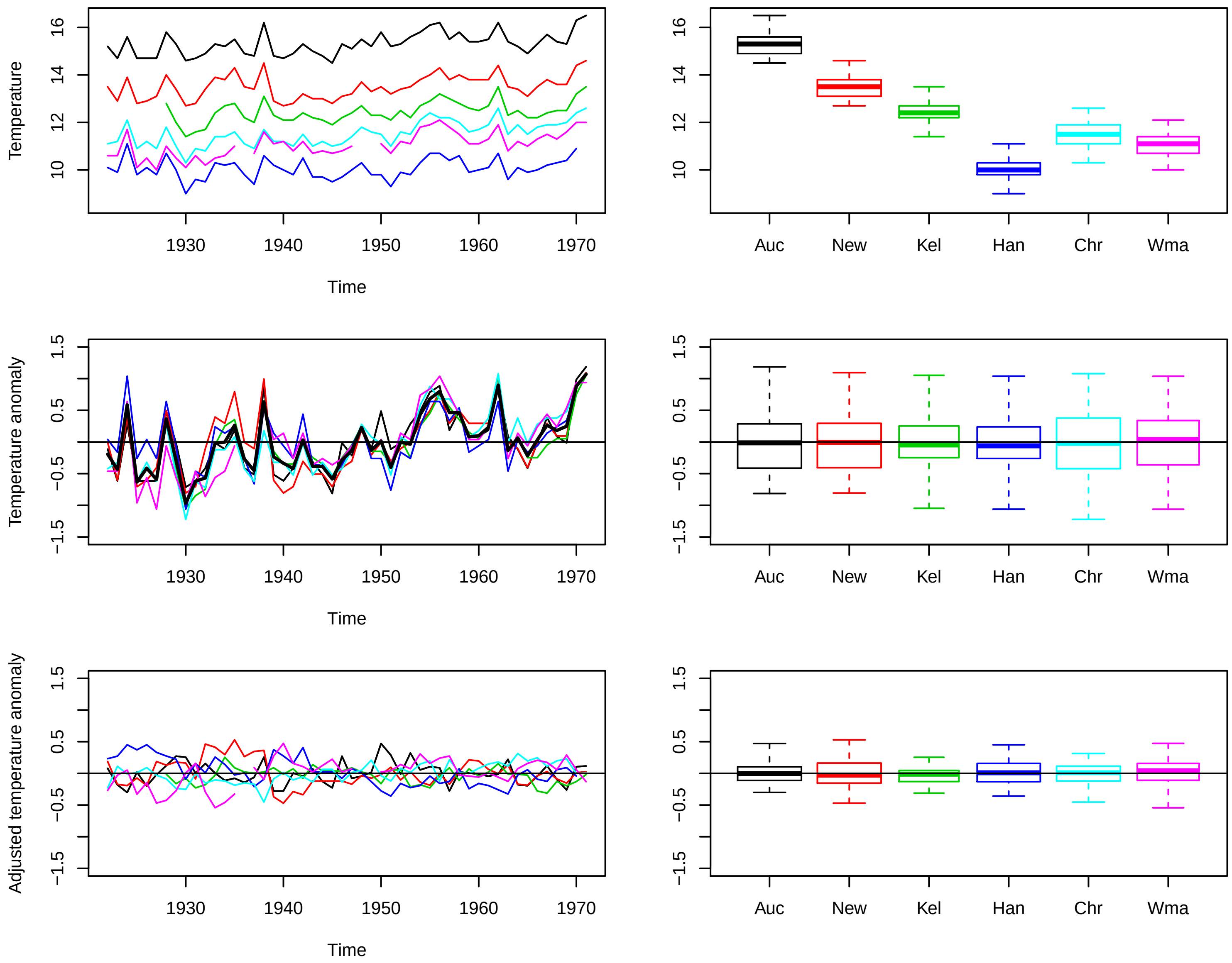
where α has been absorbed in μ_0 and the stationary errors

$$e_0(t) = X(t) + \epsilon_0(t)$$

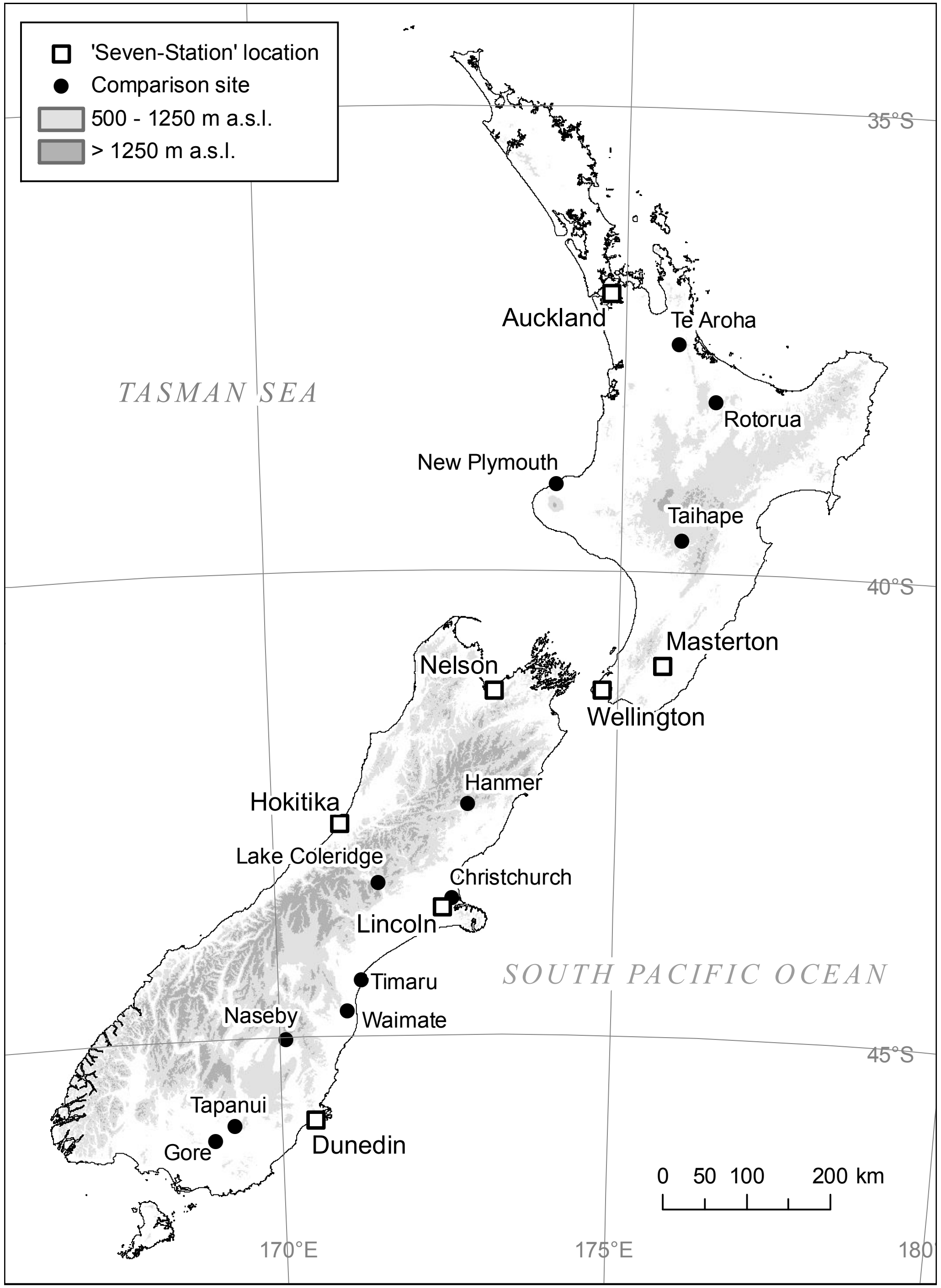
are correlated over time and space through $X(t)$.

This is a simple model which

- is widely used, in one form or another;
- approximates the true trend by a simple linear trend;
- reflects the properties of New Zealand temperatures;
- considers the simplest case of level shifts at known changepoints.



Time series and boxplots of annual average daily mean New Zealand temperatures at **Auckland**, New Plymouth, Kelburn, Hanmer Springs, Christchurch and Waimate. The **middle plots** show temperature anomalies and the **average temperature anomaly**; the bottom plots show the anomalies adjusted for average anomaly.



'Seven-Station' location
Comparison site
500 - 1250 m a.s.l.
> 1250 m a.s.l.
TASMAN SEA
SOUTH PACIFIC OCEAN
35°S
40°S
45°S
170°E
175°E
180
Auckland
Te Aroha
Rotorua
New Plymouth
Taihape
Masterton
Nelson
Wellington
Hanmer
Hokitika
Lake Coleridge
Christchurch
Lincoln
Timaru
Waimate
Naseby
Tapanui
Gore
Dunedin
0 50 100 200 km

If the δ_j were known, then the temperatures can be prior adjusted to give

$$Y_0(t) - \sum_{j=1}^{p} \delta_j d(t - t_j) = \mu_0 + \beta t + e_0(t)$$

which is a simple linear trend with stationary errors.

In practice, the δ_j are replaced by estimates $\widehat{\delta}_j$ and the model fitted to the adjusted temperatures

$$\tilde{Y}_0(t) = Y_0(t) - \sum_{j=1}^{p} \widehat{\delta}_j d(t - t_j).$$

This procedure is considered here with the δ_j estimated by local adjustment and data from nearby sites, or by fitting the model directly (global adjustment).

3. Global adjustment

Here the Site 0 model

$$Y_0(t) = \mu_0 + \beta t + \sum_{j=1}^{p} \delta_j d(t - t_j) + e_0(t)$$

is fitted directly using standard time series regression methods.

These methods include

- ordinary least squares (OLS) with standard errors adjusted for autocorrelation;

- generalised least squares (GLS) with $e_0(t)$ modelled by a suitable ARMA process

which yield linear unbiased estimators with minimum variance (GLS).

Let

$$\widehat{\beta} = \text{estimate of } \beta \text{ when the } \delta_j \text{ are unknown}$$
$$\widehat{\beta}_0 = \text{estimate of } \beta \text{ when the } \delta_j \text{ are known}$$

and consider the case of one changepoint ($p = 1$) with $e_0(t)$ white noise. Then

$$\frac{\text{Var}(\widehat{\beta}_0)}{\text{Var}(\widehat{\beta})} = 1 - \frac{3f_1(1-f_1)}{1-1/T^2}$$

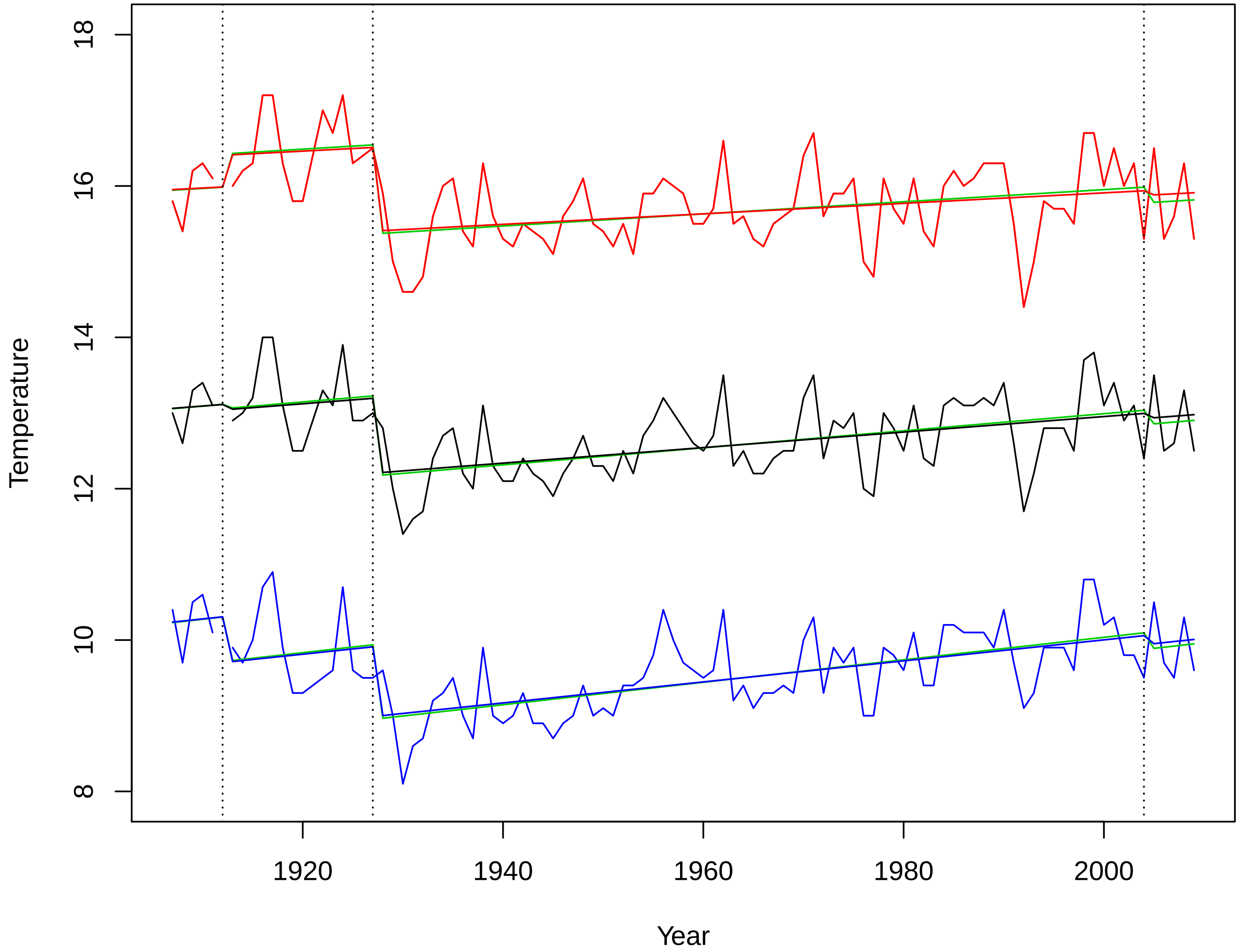
where f_1 is the fraction of observations before the changepoint.

For T large this ratio varies between 0.25 ($f_1 = 0.5$) and 1 ($f_1 = 0$ or 1). When $f_1 = 0.5$

$$\text{std dev}(\widehat{\beta}) = 2 \text{ std dev}(\widehat{\beta}_0)$$

so the presence of a changepoint can lead to a significant loss of precision.

Now apply OLS and GLS to annual average daily temperatures at Wellington over the 103 year period 1907 to 2009.



Annual average daily maximum, **mean** and minimum temperatures at Wellington. In each case a time series regression model with common slope and site specific level shifts has been fittted by GLS and the fit superimposed. The OLS fit with common slope is also shown.

Fitted OLS and GLS trends for annual average daily maximum, mean and minimum Wellington temperatures where

slope = temperature increase per century

and GLS used an MA(1) error. The OLS slope standard errors are given with and without correction for autocorrelation.

	Maximum		Mean		Minimum	
GLS	Est	SE	Est	SE	Est	SE
Slope	0.69	0.30	1.03	0.27	1.39	0.27
MA coefficient	0.32	0.09	0.30	0.10	0.30	0.10
OLS						
Slope (uncorrected)	0.80	0.24	1.13	0.22	1.49	0.22
Slope (corrected)	0.80	0.30	1.13	0.27	1.49	0.27

What improvement, if any, comes from using locally adjusted data?

4. Local adjustment

Consider a local time window about changepoint t_j (changepoint window t_j) within which

- t_j is the only Site 0 changepoint;
- temperatures at K_j neighbouring sites $Y_k(t)$ have no changepoints

and

$$Y_k(t) = \mu_k + T(t) + \epsilon_k(t)$$

where

- μ_k is the Site k temperature offset;
- $T(t)$ is the regional temperature signal as before;
- the $\epsilon_k(t)$ are mutually uncorrelated white noise processes, uncorrelated with $T(t)$ and $\epsilon_0(t)$, all with the same variance as $\epsilon_0(t)$.

Over changepoint window t_j the temperature differences

$$\begin{aligned} D_k(t) &= Y_0(t) - Y_k(t) \\ &= \mu_0 - \mu_k + \delta_j d(t - t_j) + \epsilon_0(t) - \epsilon_k(t) \end{aligned}$$

no longer involve $T(t)$. In particular the $D_k(t)$ are

- temporally independent;
- have variance $2\sigma^2$ with $\mathrm{Var}(\epsilon_k(t)) = \sigma^2$;
- have a constant correlation of 0.5 across sites.

Now estimate δ_j by linear estimators of the form

$$\hat{\delta}_j = \sum_{t=1}^{T} \sum_{k=1}^{K} w_k^{(j)}(t) D_k(t)$$

where K is the total number of comparison sites and the known weights $w_k^{(j)}(t)$ are zero for times t and sites k unrelated to changepoint window t_j.

5. Fitting a linear trend to adjusted data

A natural estimator of β commonly used in practice is

$$\tilde{\beta} = \sum_{t=1}^{T} u_0(t)\tilde{Y}_0(t) = \sum_{t=1}^{T-1} v_0(t)\Delta\tilde{Y}_0(t+1)$$

where

$$\tilde{Y}_0(t) = Y_0(t) - \sum_{j=1}^{p} \widehat{\delta}_j d(t - t_j), \qquad v_0(t) = \sum_{s=t+1}^{T} u_0(s) = -\sum_{s=1}^{t} u_0(s)$$

are the locally adjusted data and the $u_0(t)$ are the OLS or GLS weights that would be used in the case of known changepoints.

What are the statistical properties of $\tilde{\beta}$?

First note that $\tilde{\beta}$ can be written as

$$\tilde{\beta} = \widehat{\beta}_0 + \sum_{j=1}^{p} v_0(t_j)(\widehat{\delta}_j - \delta_j)$$

where $\widehat{\beta}_0$ denotes the estimator of β when the δ_j are known.

This gives

$$E(\tilde{\beta}) = \beta$$

and

$$\text{Var}(\tilde{\beta}) = \text{Var}(\hat{\beta}_0) + 2\sum_{j=1}^{p} v_0(t_j)\text{cov}(\hat{\beta}_0, \hat{\delta}_j) + \sum_{i=1}^{p}\sum_{j=1}^{p} v_0(t_i)v_0(t_j)\text{cov}(\hat{\delta}_i, \hat{\delta}_j)$$

where

$$\text{cov}(\hat{\beta}_0, \hat{\delta}_j) = \sigma^2 \sum_{t=1}^{T} u_0(t) w^{(j)}(t)$$

$$\text{cov}(\hat{\delta}_i, \hat{\delta}_j) = \sigma^2 \sum_{t=1}^{T} (w^{(ij)}(t) + w^{(i)}(t) w^{(j)}(t))$$

and

$$w^{(ij)}(t) = \sum_{k=1}^{K} w_k^{(i)}(t) w_k^{(j)}(t), \quad w^{(i)}(t) = \sum_{k=1}^{K} w_k^{(i)}(t).$$

All terms on the right-hand side are readily computed with

- $\text{Var}(\hat{\beta}_0)$ a function of $\text{Var}(Y_0(t)) = \text{Var}(T(t)) + \sigma^2$;
- the other terms proportional to the measurement error σ^2 only.

Restrict attention to estimating the level shifts δ_j by

$$\hat{\delta}_j = \frac{1}{K_j} \sum_k (\bar{D}_k^-(t_j) - \bar{D}_k^+(t_j))$$

where the sum is over the K_j comparison sites for changepoint window t_j and

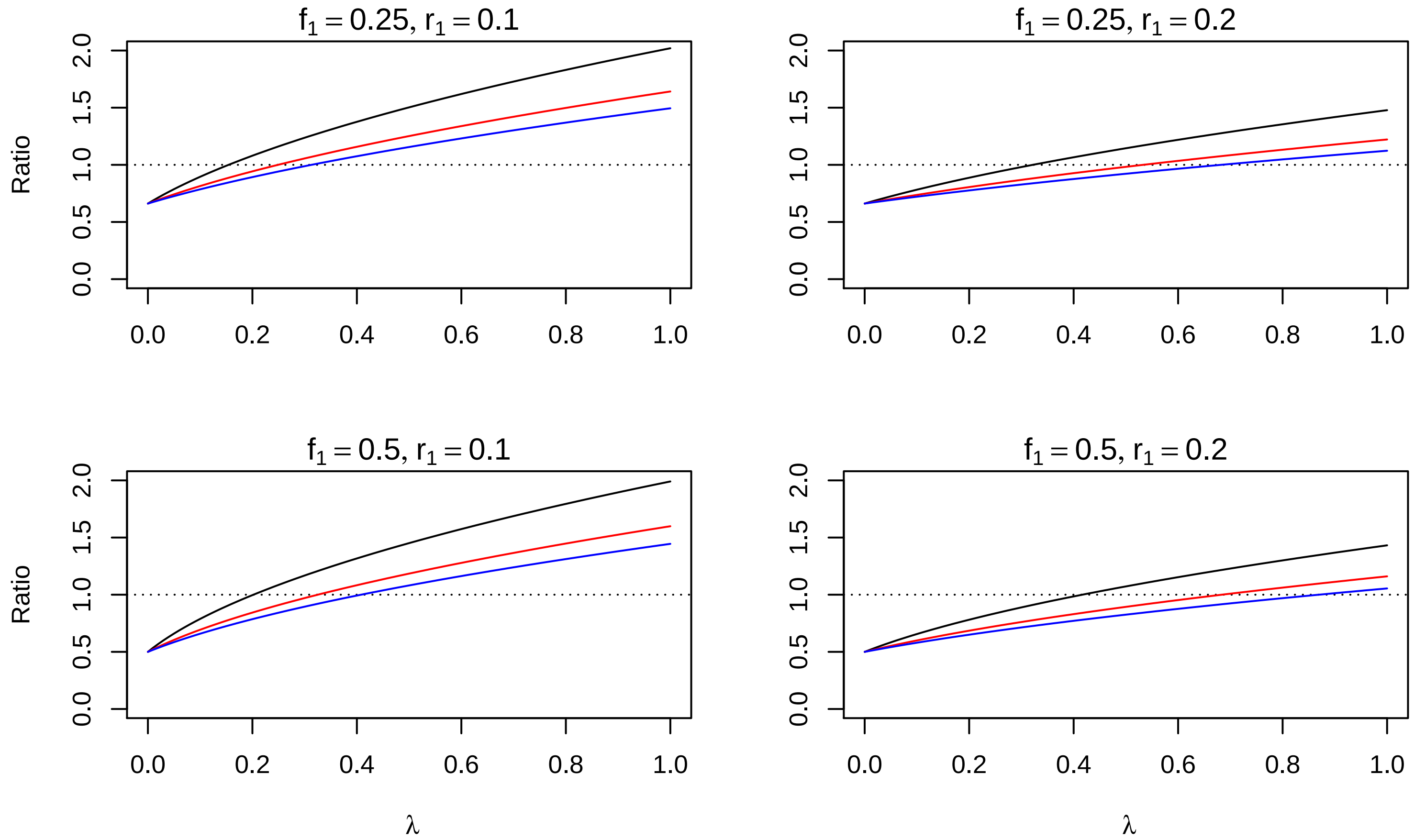
$$\bar{D}_k^-(t_j) = \text{mean of the } n_j^- \text{ temperature differences } D_k(t) \text{ up to } t_j$$
$$\bar{D}_k^+(t_j) = \text{mean of the } n_j^+ \text{ temperature differences } D_k(t) \text{ after } t_j$$

Commonly used estimator with simple weights $w_k^{(i)}(t)$.

Local or global adjustment: which gives the best estimate of β?

Consider the simple case where

- there are no missing values;
- there is only one changepoint centrally located in the changepoint window;
- the errors $e_0(t)$ are white noise;
- $\lambda = \sigma^2/\text{Var}(e_0(t)) = \text{Var}(\epsilon_k(t))/\text{Var}(Y_0(t))$.



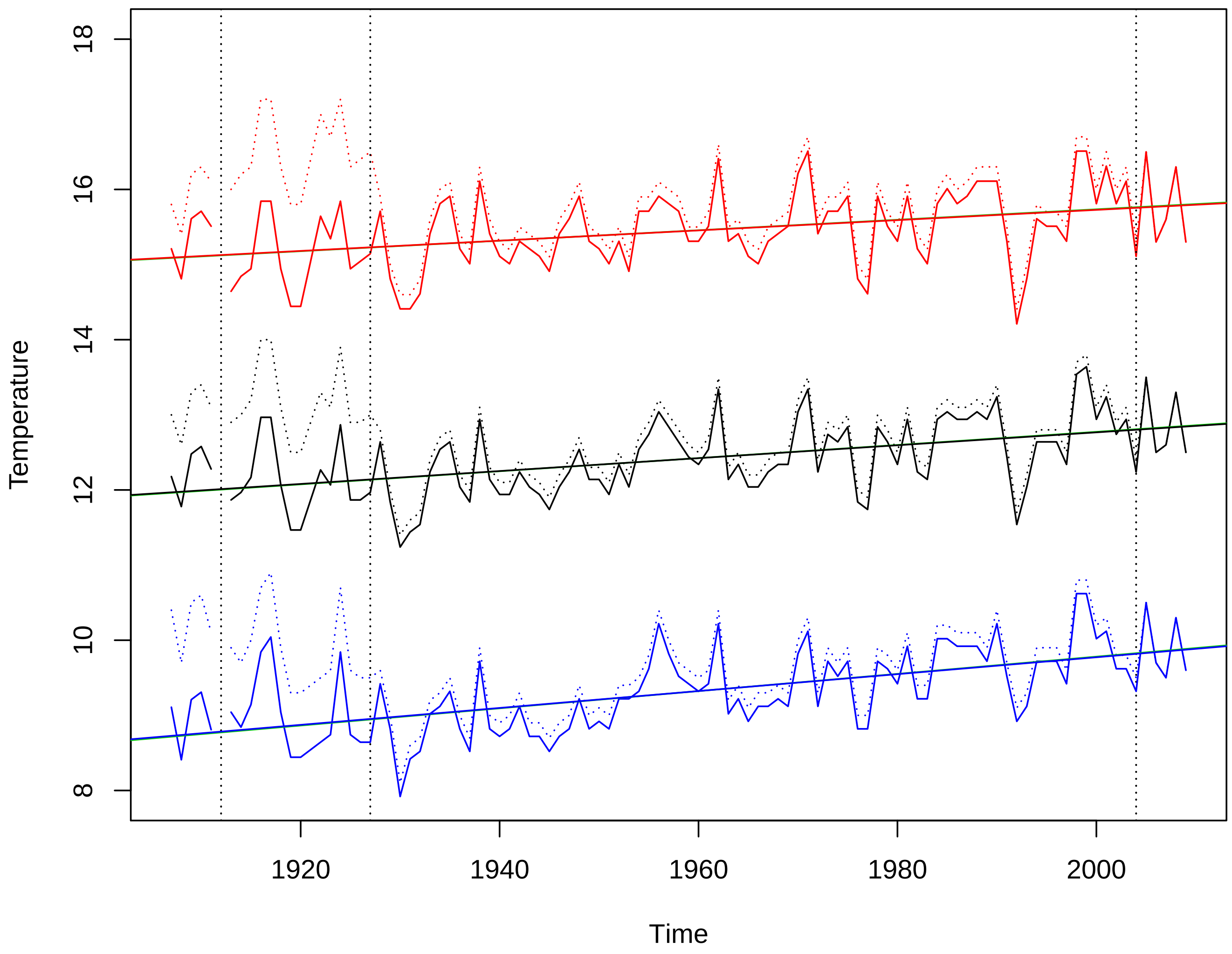
Ratio of the standard error of the slope estimate $\tilde{\beta}$ for locally adjusted temperatures to that of the GLS slope estimate $\widehat{\beta}$ as a function of λ. Assume a central changepoint at $f_1 T$ in a window of width $r_1 T$ and white noise errors. Here T is the series length and the number of comparison stations is $K_1 = 1$, $K_1 = 4$ and $K_1 = \infty$.

Now apply these results to annual average daily temperatures at Wellington over the 103 year period 1907 to 2009.

Key design parameters for the Wellington site changes are

	Buckle Street	Thorndon	Kelburn
Changepoint t_j	1912	1927	2004
Changepoint window	1907–1922	1918–1937	1995–2009
Window length	16	20	15
Number of missing values in window	10	4	0
Number of comparison stations K_j	3	4	1
Comparison stations	Auc, Chr, Nel	Auc, Chr, Tai, Wai	Chr

The comparison stations are Auckland (Auc), Christchurch (Chr), Nelson (Nel), Taihape (Tai) and Waingawa (Wai).



Locally adjusted annual average daily maximum, **mean** and minimum temperatures at Wellington. Linear time trends have been fitted by GLS with the OLS fits superimposed. Original data (dotted) also shown for reference.

Estimates of the rate of increase per 100 years at Wellington, New Zealand, for annual average daily maximum, mean and minimum temperatures.

	Maximum		Mean		Minimum	
Local adjustment	Est	SE	Est	SE	Est	SE
OLS: no corrections	0.70	0.16	0.87	0.15	1.15	0.15
OLS: corrected for autocorrelation	0.70	0.20	0.87	0.18	1.15	0.19
OLS: both corrections	0.70	0.24	0.87	0.21	1.15	0.22
GLS: no homogenisation correction	0.68	0.20	0.86	0.18	1.12	0.18
GLS: corrected for homogenisation	0.68	0.24	0.86	0.20	1.12	0.22
Global adjustment						
OLS: no autocorrelation correction	0.80	0.24	1.13	0.22	1.49	0.22
OLS: corrected for autocorrelation	0.80	0.30	1.13	0.27	1.49	0.27
GLS	0.69	0.30	1.03	0.27	1.39	0.27

Note that

- OLS and GLS slope estimates are similar;
- standard errors of the slope estimates for local adjustment are around 20% lower than those for global adjustment.

Local adjustment helps!

6. Conclusions

- OLS and GLS slope estimates are generally consistent and unbiased.
- Failure to account for local adjustment (homogenisation) can lead to slope estimates with standard errors that are biased downwards (in this case around 15%; up to 30% if OLS without autocorrelation correction used).
- Slope estimates from locally adjusted data should be more accurate than estimates from single-site global regression models.

Future work includes

- extension to monthly data and seasonality;
- allowance for overlapping records;
- other forms of homogenisation.