Stockholms
universitet

Multiple Measurement Error Regression with autocorrelated errors in predictors as a prediction method

Ekaterina Fetisova

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2012:8
Matematisk statistik
September 2012

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2012:8**
http://www.math.su.se

Multiple Measurement Error Regression with autocorrelated errors in predictors as a prediction method

Ekaterina Fetisova*

September 2012

Abstract

The aim of this master thesis is to evaluate the performance of multiple linear functional measurement error (ME) regression with autocorrelated errors in predictors as a prediction method. Setting it in a climate context, the aim is to investigate whether this method can be used for predictions of past values of temperature over a large region. Using the primary data representing true values a large number of datasets were generated by the model of interest. Because multicollinearity was not detected all five given true predictors have been included in the model. To achieve independency in the errors an appropriate transformation has been applied. Based on Monte-Carlo methods, the results have indicated that data do not support consistent parameter estimation under ME model with no error in the equation independently of how strong autocorrelation is and how large the error variation is. The same problem was present even under the ME model with an error in the equation though not for data where the error variation accounted for 20% of the total variation in each observed predictor (which was the lowest level of the error variation analysed). Using this type of data the model has demonstrated an adequate prediction performance in terms of MSEP. However the long run analysis of confidence intervals (CI's) with the nominal confidence level 0.9 has indicated a large variability in possible values of the actual coverage probability, suggesting to use this model as a prediction method with great caution even if the error variation is modest. Further the thesis aims to illustrate the inappropriateness of the use of models, which either do not take into consideration autocorrelation in measurement errors or do not allow for errors in predictors at all, when data contain predictors with autocorrelated errors. Based on the same original datasets above, the analysis has indicated the high probability of obtaining of extremely large estimators under ME regression, assuming uncorrelated errors in each predictor, both with no error in the equation and with an error in the equation. The inappropriateness of ordinary multiple regression, whose estimators turned out to take reasonable values, has been detected under the long run analysis of 90% CI's: the estimated coverage probabilities turned out to be less than 0.45 for all magnitudes of the error variation.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: katarina@math.su.se. Supervisor: Gudrun Brattström.

Foreword and acknowledgements

The present analysis has been performed at Stockholm university at department of Mathematics, division of Mathematical Statistics in collaboration with Bert Bolin Centre for Climate Research. I would like to sincerely thank my supervisors, Gudrun Brattström at department of Mathematics and Anders Moberg at Bert Bolin Centre, for all worthwhile discussions we had and all correspondence during this work. I would also like to thank Eduardo Zorita for his help with providing data for the analysis.

Contents

1 Background

In order to understand the current and future climate variations it is of great importance to have a knowledge about the climate variation in the past. Unfortunately, systematic instrumental measurements of different climate variables such as precipitation, drought or near-surface temperature are mostly availablfe only for the past 130-150 years, which corresponds to so called industrial period. For getting an idea of pre-industrial climates, data from various climate proxy 'archives' like historical documents, annual growth rings in trees, sediments from lake bottoms and so on are used. In contrast to instrumental data, proxies are available for a long time period, e.g. 1000 years. Usually data from the industrial period, when both instrumental data and proxies are available, are used for estimation of a statistical relationship between a climate variable of interest and proxies, from which the name "calibration period" arises. Proxy data from the pre-industrial period together with estimated parameters are used for reconstructions of past values of that climate variable, from which the name "reconstruction period" comes. However, the procedure of estimation and reconstruction is not a clear-cut process and it has already caused numerous discussions, showing the importance of collaboration between palaeoclimatologists and statisticians.

2 Introduction

The idea about this project arose during the discussions I had with my supervisors Gudrun Brattström and Anders Moberg in December 2011 - January 2012. One of the main concerns for the modern palaeoclimatology is the developing of a reliable statistical method for reconstructions of the climate of the past. Numerous reconstruction methods were suggested by different researches depending on (1) what assumptions about errors in climate proxy and instrumental data were made and (2) the spatial characteristics of the data. One of the methods, suggested by G. Brattström and A. Moberg ([9]), is a univariate functional measurement error model that allows for white noise in instrumental data (namely temperature) and autocorrelated errors in proxy variable measured at a certain location. The presence of only a single proxy variable made it possible to apply the method of calibration for reconstruction of temperature, which implies initially regression of a single proxy record on a single instrumental record. Naturally, this analysis has excited interest in further analysis of measurement error model allowing climate reconstructions over a larger region by using several proxy records from multiple sites.

3 Aim

From the palaeoclimate viewpoint, the aim is to investigate whether ME models can be used as a reliable statistical method for reconstructions of temperature of the past over a large region, given proxies from different locations within this region and one sample of mean temperature over the same region. Furthermore, the model should take into account that proxy data are contaminated with positive autocorrelated measurement errors and instrumental data (temperature) are contaminated with uncorrelated errors. The same assumptions were assumed even in ([9]), although it can be a simplification of a real character of measurement errors in instrumental data ([9], p.320). As several variables representing proxies are involved, multiple ME regression of instrumental data on proxy data has been suggested. So from the statistical point of view, the aim is to evaluate the performance of multiple measurement error regression allowing for autocorrelated errors in predictors as a prediction method. Two types of the model will be analysed - with no error in the equation and with an error in the equation. In addition, it is assumed that the true values of all variables are constants, which means that the ME functional model should be analysed.

Further inspired by analyses performed by some other researches the thesis aims to illustrate the inappropriateness of the use of models, which do not take into account either the presence of measurement errors or autocorrelation in errors, when data, they are fitted to, contain predictors with autocorrelated errors. The models are: ordinary multiple regression, assuming fixed predictors and ME regressions, both with no error in the equation and with an error in the equation, assuming uncorrelated errors in each predictor.

4 Method for model validation

In this analysis the method for model validation is based on the idea of cross-validation. According to this idea, a given dataset is divided into two subsets in order to use one of them for model fitting and the other for prediction and assessment of accuracy of prediction. In this work prediction is associated with regression model which, as known, does not establish a causal connection between variables. It is defined as a new past value of temperature calculated from the regression model and new past values of predictors (proxies). For assessment of predictive abilities of the models, or accuracy of prediction, the known predictand will be compared with the predicted one by means of Mean Squared Error of Prediction, MSEP, (see section 6.3).

Data used in this analysis were generated in such a way that the assumptions about error terms described in section Aim were satisfied. By simulating error terms repeatedly a large number of sets of observed data can be formed (see details in section Description of data). This enables one to apply Monte-Carlo methods in order to perform a long run analysis of predictive abilities of the models.

5 Theory

5.1 Multiple Measurement error model with no error in the equation

The functional model with a single true response variable y_t and the vector of true values $\mathbf{x}_t$ is defined as follows ([6], ch. 2.3):

$$\begin{aligned} y_t &= \boldsymbol{x}_t\boldsymbol{\beta}, \\ (Y_t, \boldsymbol{X}_t) &= (y_t, \boldsymbol{x}_t) + (\epsilon_t, \boldsymbol{u}_t), \end{aligned} \tag{5.1.1}$$

$$Y_t = \boldsymbol{x}_t\boldsymbol{\beta} + \epsilon_t$$

for $t = 1, 2, \ldots, n$,

where $\{\boldsymbol{x}_t\}$ is a fixed k-dimensional row vector

$(Y_t, \boldsymbol{X}_t)$ are observed values

$\boldsymbol{a}_t = (\epsilon_t, \boldsymbol{u}_t)'$ is the vector of measurement errors,
such that $\boldsymbol{a}_t' \sim NI(\boldsymbol{0}, \boldsymbol{\Sigma}_{aa})$.

Let $\boldsymbol{\Sigma_{aa}}$ to be known. Then the maximum likelihood estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{M_{XX}} - \hat{\lambda}\boldsymbol{\Sigma_{uu}})^{-1}(\boldsymbol{M_{XY}} - \hat{\lambda}\boldsymbol{\Sigma_{u\epsilon}}) \tag{5.1.2}$$

provided $\boldsymbol{M_{XX}} - \hat{\lambda}\boldsymbol{\Sigma_{uu}}$ is nonsingular, where $\boldsymbol{M_{XX}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_t'\boldsymbol{X}_t$ and $\hat{\lambda}$ is the smallest root of

$$\Big|\boldsymbol{M_{ZZ}} - \lambda\boldsymbol{\Sigma_{aa}}\Big| = 0,$$

where $\boldsymbol{Z} = (Y_t, \boldsymbol{X}_t)$.

Furthemore, if $n \geq p$ and $|\boldsymbol{M}_{xx}| > 0$, then

$$\boldsymbol{\Gamma}_{\beta\beta}^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{I}) \tag{5.1.3}$$

as $n \to \infty$ with covariance matrix

$$\boldsymbol{\Gamma}_{\beta\beta} = n^{-1}\Big(\boldsymbol{M}_{xx}^{-1}\sigma_{\nu\nu} + \boldsymbol{M}_{xx}^{-1}(\boldsymbol{\Sigma}_{uu}\sigma_{\nu\nu} - \boldsymbol{\Sigma}_{u\nu}\boldsymbol{\Sigma}_{\nu u})\boldsymbol{M}_{xx}^{-1}\Big), \tag{5.1.4}$$

where $\nu_t = \epsilon_t - \boldsymbol{u}_t\boldsymbol{\beta}$, $\sigma_{\nu\nu} = (1, -\boldsymbol{\beta}')\boldsymbol{\Sigma_{aa}}(1, -\boldsymbol{\beta}')'$, and $\Sigma_{\boldsymbol{u}\nu} = \Sigma_{\boldsymbol{u}\epsilon} - \Sigma_{\boldsymbol{uu}}\boldsymbol{\beta}$

Remark: if $\mathrm{Var}(\epsilon_t) = \sigma_{\epsilon\epsilon}$ is unknown, the method of maximum likelihood fails to yield consistent estimators for all parameters of the model. This illustrates the problem of model identifiability. According to ([3], p.5) if $\boldsymbol{Z}$ is a random vector whose distribution is from some family $\mathcal{F} = \{\mathcal{F}_\theta; \boldsymbol{\theta} \in \Theta\}$, then the parameter, θ_i, the ith component of the vector, $\boldsymbol{\theta}$, is identifiable if and only if no two values of $\boldsymbol{\theta} \in \Theta$, whose ith components differ, lead to the same distribution of $\boldsymbol{Z}$. The vector of parameters, $\boldsymbol{\theta}$, is said to be identifiable if and only if all its components are identifiable. The model is said to be identifiable if $\boldsymbol{\theta}$ is identifiable. The typical identifiability assumption for ME model with no error in the equation is that the entire error covariance structure, including $\sigma_{\epsilon\epsilon}$, is known or is known up to a scalar multiple ([6], [3]). In practice, one often has an estimator of $\Sigma_{\boldsymbol{aa}}$. Oddly enough, identifiability is not sufficient for consistent estimation ([3], p.239). Other additional conditions should be available to support it. For ME model with no error in the equation such a condition is a positive definite denominator in the estimator of $\boldsymbol{\beta}$ (for the definition of consistent estimators see [8], p.270).

5.2 Multiple Measurement error model with an error in the equation

If the true values y_t and $\boldsymbol{x}_t$ are not perfectly related, in other words if there are factors other than $\boldsymbol{x}_t$ that are responsible for variation in y_t, one might specify

([6], p.106-108)

$$y_t = \boldsymbol{x}_t\boldsymbol{\beta} + q_t$$
$$(Y_t, \boldsymbol{X}_t) = (y_t, \boldsymbol{x}_t) + (\epsilon_t, \boldsymbol{u}_t), \tag{5.2.1}$$
$$Y_t = \boldsymbol{x}_t\boldsymbol{\beta} + e_t$$

where the q_t are independent $(0, \sigma_{qq})$ random variables, and q_t is independent of $\boldsymbol{x}_j$ for all t and j. The random variable q_t is called the *error in the equation*. $(\epsilon_t, \boldsymbol{u}_t)' = \boldsymbol{a}_t'$ is a vector of measurement errors, $\boldsymbol{a}_t' \sim \mathrm{NI}(\boldsymbol{0}, \Sigma_{\boldsymbol{aa}})$, and $\boldsymbol{a}_t'$ is independent of $(q_j, \boldsymbol{x}_j)$ for all t and j. Further, the random variable e_t is the sum of an error made in measuring y_t and an error in the equation, $e_t = \epsilon_t + q_t$. Let $\Sigma_{\boldsymbol{aa}}$ to be known. Then the estimator of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}} = (\boldsymbol{M}_{XX} - \Sigma_{\boldsymbol{uu}})^{-1}(\boldsymbol{M}_{XY} - \Sigma_{\boldsymbol{u}\epsilon})$$

provided $\boldsymbol{M}_{XX} - \Sigma_{\boldsymbol{uu}}$ is nonsingular, and a consistent estimator of σ_{qq} is

$$\hat{\sigma}_{qq} = \hat{\sigma}_{\nu\nu} - (\sigma_{\epsilon\epsilon} - 2\hat{\boldsymbol{\beta}}\Sigma_{\boldsymbol{u}\epsilon} + \hat{\boldsymbol{\beta}}\Sigma_{\boldsymbol{uu}}\hat{\boldsymbol{\beta}}),$$

where $\nu_t = e_t - \boldsymbol{u}_t\boldsymbol{\beta},$, $\sigma_{\nu\nu} = (n-k)^{-1}\sum_{t=1}^{n}(Y_t - \boldsymbol{X}_t\boldsymbol{\beta})^2$.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma_{qq})'$ and let $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}', \tilde{\sigma}_{qq})'$. If $n \geq p$ and $\boldsymbol{M}_{xx}$ is positive definite, then

$$\boldsymbol{\Gamma}^{-1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{I}),$$

where the submatrices of $\boldsymbol{\Gamma}$ are

$$\boldsymbol{\Gamma}_{\beta\beta} = n^{-1}\Big(\boldsymbol{M}_{xx}^{-1}\sigma_{\nu\nu} + \boldsymbol{M}_{xx}^{-1}[\boldsymbol{\Sigma}_{uu}\sigma_{\nu\nu} + \boldsymbol{\Sigma}_{u\nu}\boldsymbol{\Sigma}_{\nu u}]\boldsymbol{M}_{xx}^{-1}\Big),$$

$$\Gamma_{qq} = n^{-1}\mathrm{Var}(\nu^2)$$
$$\Gamma_{\beta q} = 2n^{-1}\boldsymbol{M}_{xx}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{u}\nu}\sigma_{\nu\nu}$$

where $\Sigma_{\boldsymbol{u}v} = \Sigma_{\boldsymbol{u}\epsilon} - \Sigma_{\boldsymbol{uu}}\boldsymbol{\beta}$.

5.3 Univariate autoregressive process of order 1

The first order autoregressive process, AR(1), is given by ([1], p.17-18)

$$X_t = \rho X_{t-1} + a_t, \quad t = 0, \pm 1, \pm 2, \ldots, \tag{5.3.1}$$

where $\{a_t\} \sim \mathrm{WN}(0, \sigma_a^2)$, $|\rho| < 1$, and a_t is uncorrelated with X_s for each $s < t$. It can be showed that $E[X_t] = 0$ and $\mathrm{Var}(X_t) = \frac{\sigma_a^2}{1-\rho^2}$ for all t.

5.4 Other definitions

5.4.1 White Noise Processes

A process $\{a_t\}$ is called a white noise process if it is a sequence of uncorrelated random variables, each with zero mean and variance σ^2. This is indicated by the notation ([1], ch.1)

$$\{a_t\} \sim \mathrm{WN}(0, \sigma^2).$$

5.4.2 Gaussian Filter Coefficients

Let σ_f be a nonnegative integer. By applying the Gaussian filter to a given unsmoothed time series, X_t, $t = 1, 2, \dots, n$, a smoothed time series

$$X_t^{(w)} = \sum_{i=-2\sigma_f}^{2\sigma_f} w_i X_{t-i}, \qquad 2\sigma_f + 1 \le t \le n - 2\sigma_f,$$

is obtained, where the Gaussian filter coefficients are given by ([9])

$$w_i = \frac{v_i}{\sum_{j=-2\sigma_f}^{2\sigma_f} v_j},$$

where

$$v_i = exp\{-i^2/(2 \cdot \sigma_f^2)\}, \quad i = -2\sigma_f, -2\sigma_f + 1, \dots, 2\sigma_f.$$

For $\sigma_f = 0$ the filter has only one term, $w = 1$, implying an unsmoothed series. Palaeoclimatologists use often this filter in order to study variations at larger temporal scales than one year, provided the seasonality is not present. The usual values of σ_f in palaeoclimatology are 3 and 9. Only $\sigma_f = 9$ is used in this study.

The gaussian filter coefficients have the following properties:

$$w_i = w_{-i}, \quad i \le |2\sigma_f|$$

$$\sum_{i=-2\sigma_f}^{2\cdot\sigma_f} w_i = 1, \qquad \sum_{i=-2\sigma_f}^{2\cdot\sigma_f} w_i^2 \le 1,$$

$$\sum_{i=-2\sigma_f}^{2\cdot\sigma_f} w_i^2 \to 0 \quad \text{as} \quad \sigma_f \to \infty.$$

As implied by its name, the Gaussian filter coefficients form a bell-shaped curve, which is characteristic for the Gaussian (normal) distribution. The following figure shows the Gaussian filter coefficients for $\sigma_f = 9$. Only $\sigma_f = 9$ is used in this study.

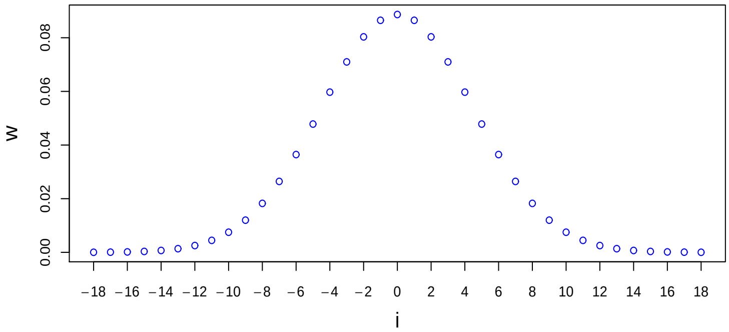


Figure 1. Gaussian filters coefficients for $\sigma_f = 9$.

6 Statistical analysis

6.1 Description of data

Data for the present analysis were formed according to an approach known among palaeoclimatologists as a *pseudo-proxy experiment* (PPE, [12]). This approach, firstly, reduces substantially the number of the sources of uncertainty, allowing researches to evaluate the prediction performance of a statistical method of interest for a certainty. Secondly, it provides a longer validation period than that which can be achieved with real-world instrumental data. Lastly, the PPE allows one to "observe" data that are absolutely unobservable in reality, e.g. temperature in the Middle Ages. According to the PPE observed data are formed by adding simulated error terms with desired properties to the artificially produced primary data, representing true values. By simulation of error terms repeatedly, a large number of sets of observed data can be formed. In this analysis each set of observed data contains 490 observations of six variables:

$$(\underbrace{Y, Z_1, Z_2, Z_3, Z_4, Z_5}_{\text{Observed variables}}) =$$

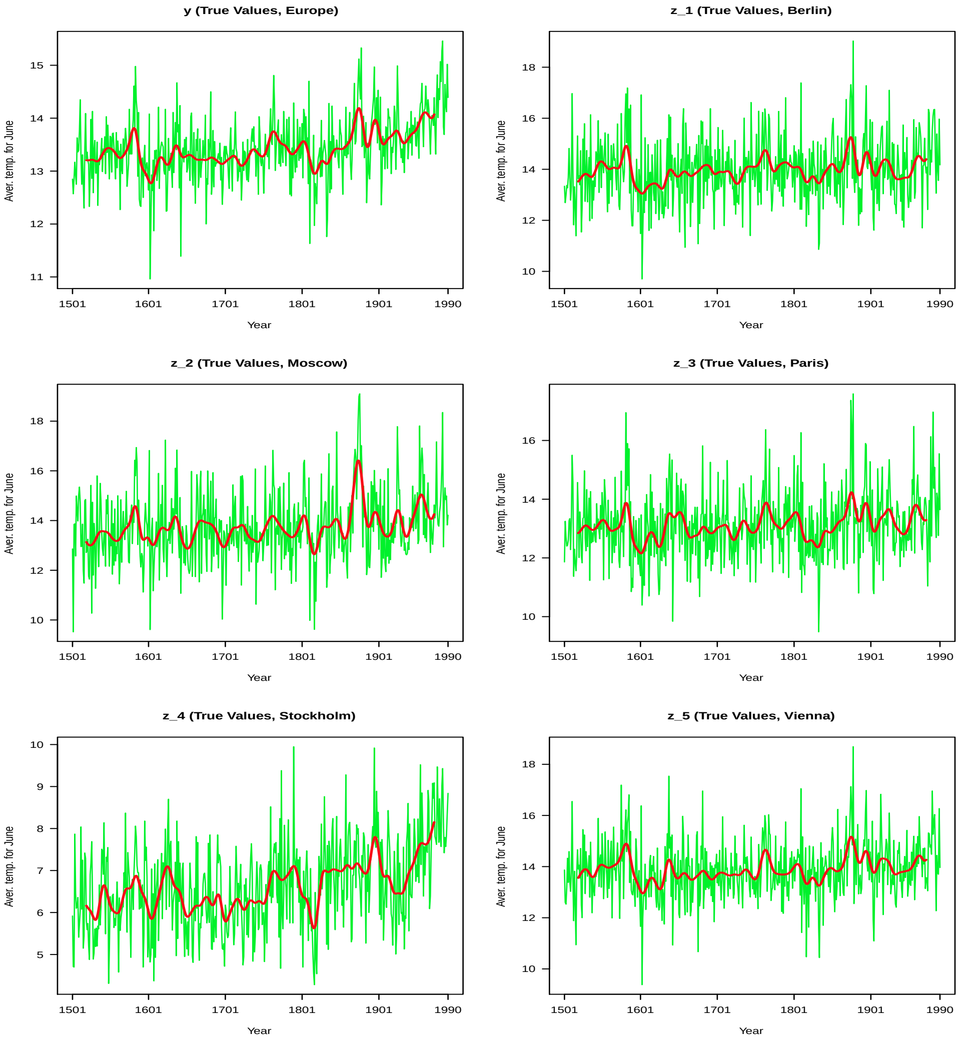
$$= (\underbrace{y, z_1, z_2, z_3, z_4, z_5}_{\text{True values} \quad \text{(given)}}) + (\underbrace{\epsilon, u_1, u_2, u_3, u_4, u_5}_{\text{Measurement errors}})$$

For the purpose of this analysis, $\{z_i\}$ for all i will represent the true predictors, that implies directly that $(Z_1, Z_2, Z_3, Z_4, Z_5)$ will represent the observed predictors. Thus the true and observed predictand will be represented by $\{y\}$ and $\{Y\}$ respectively.

6.1.1 True values

All samples of the primary data, representing true values, were selected from a certain climate model [1] [7], which is a complex system developed specially for simulations of climate at different spatial scales with respect to different sources of external forcings like greenhouse gas concentrations in the atmosphere, solar irradiance and volcanic aerosols. The given true values represent the June mean temperature from 1501 to 1990 in Berlin (z_1), Moscow (z_2), Paris (z_3), Stockholm (z_4), Vienna (z_5) and the June mean temperature averaged over the whole of Europe (y). In Figure 2 all these series are shown. It appears that they are stationary in the variance, but probably not in the mean, especially $\{y_t\}$ and $\{z_{t,4}\}$. Physically, this is essentially due to the influence from the temporally increasing amount of atmospheric greenhouse gases applied in the simulation. The analysis of the sample autocorrelation functions and the sample partial autocorrelation functions with the following application of the Dickey-Fuller test ([1], p.194; [11], chapter 5) has indicated that all series are stationary around their deterministic trend. Fitting the polynomials of different orders to each time series has shown that either a linear or quadratic trend, depending on the series, is statistically significant.

[1] The Regional Circulation Model (RCM) MM5 coupled to the Global Circulation Model (GCM) ECHO-G



`Figure 2.` *The unsmoothed original true values in green. The smoothed with the Gaussian filter original true values in red* ($\sigma_f = 9$).

In Figure 3 the detrended time series are plotted (hereafter $\{y_t\}, \{z_{t,1}\}, \{z_{t,2}\}, \{z_{t,3}\}, \{z_{t,4}\}, \{z_{t,5}\}$ are referred to as detrended true series). Apparently, after detrending all series resemble realizations of stationary time series. Their sample variances (at lag 0) are equal to

$$\widehat{\mathrm{Var}}(y), \widehat{\mathrm{Var}}(z_1), \widehat{\mathrm{Var}}(z_2), \widehat{\mathrm{Var}}(z_3), \widehat{\mathrm{Var}}(z_4), \widehat{\mathrm{Var}}(z_5) =$$

$$= (0.3017, \quad 1.6203, \quad 2.0073, \quad 1.4068, \quad 0.9360, \quad 1.3216)$$

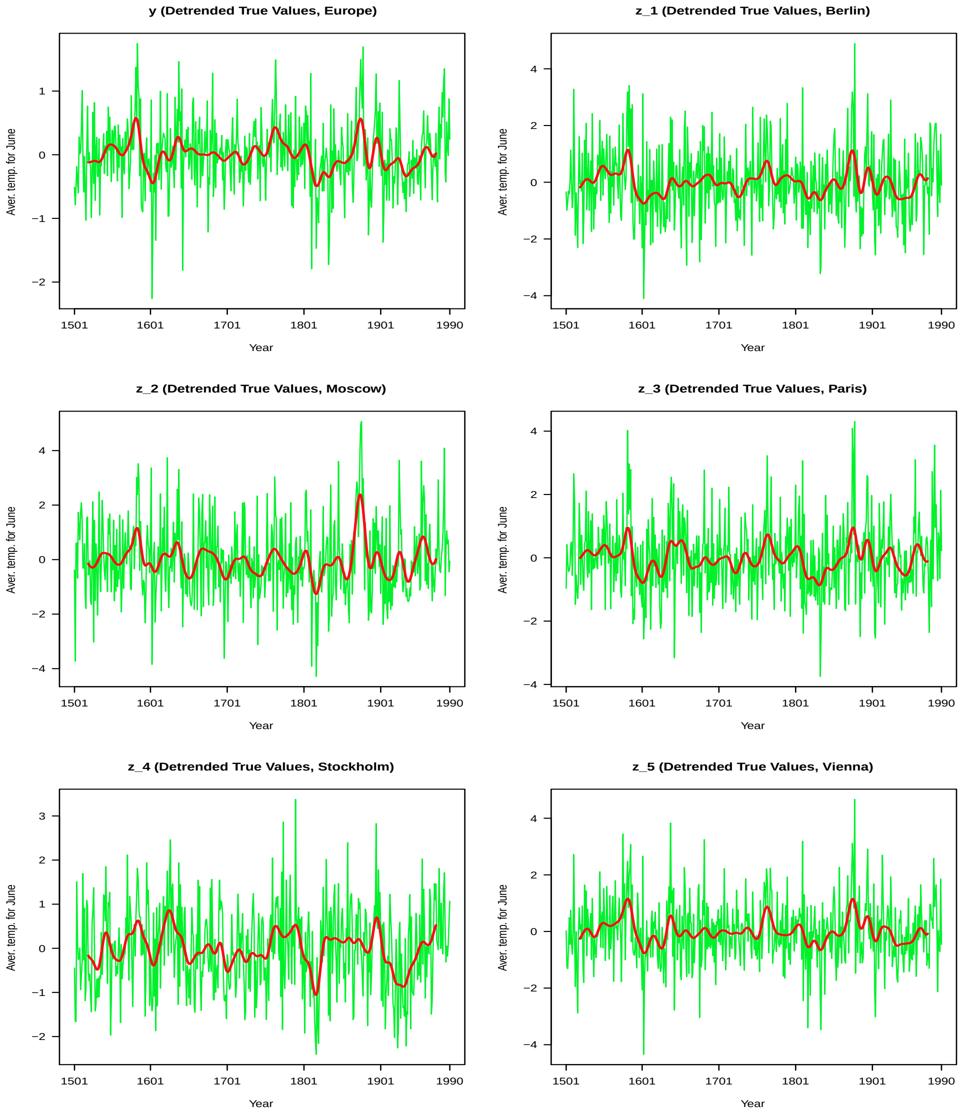


Figure 3. *The unsmoothed detrended true values in green. The smoothed with the gaussian filter detrended true values in red* ($\sigma_f = 9$).

The scatterplots in Figure 4 show all the correlations (crosscorrelations at lag 0) between the detrended true series. As one can see the correlations are positive, which is expected, with the highest correlation between Berlin (z_1) and Vienna (z_5) that is equal to 0.842. Furthermore, the 5 scatterplots in the upper row confirm the expectation of a linear relationship between true values of the observed predictand and true values of the observed predictors over the whole period.

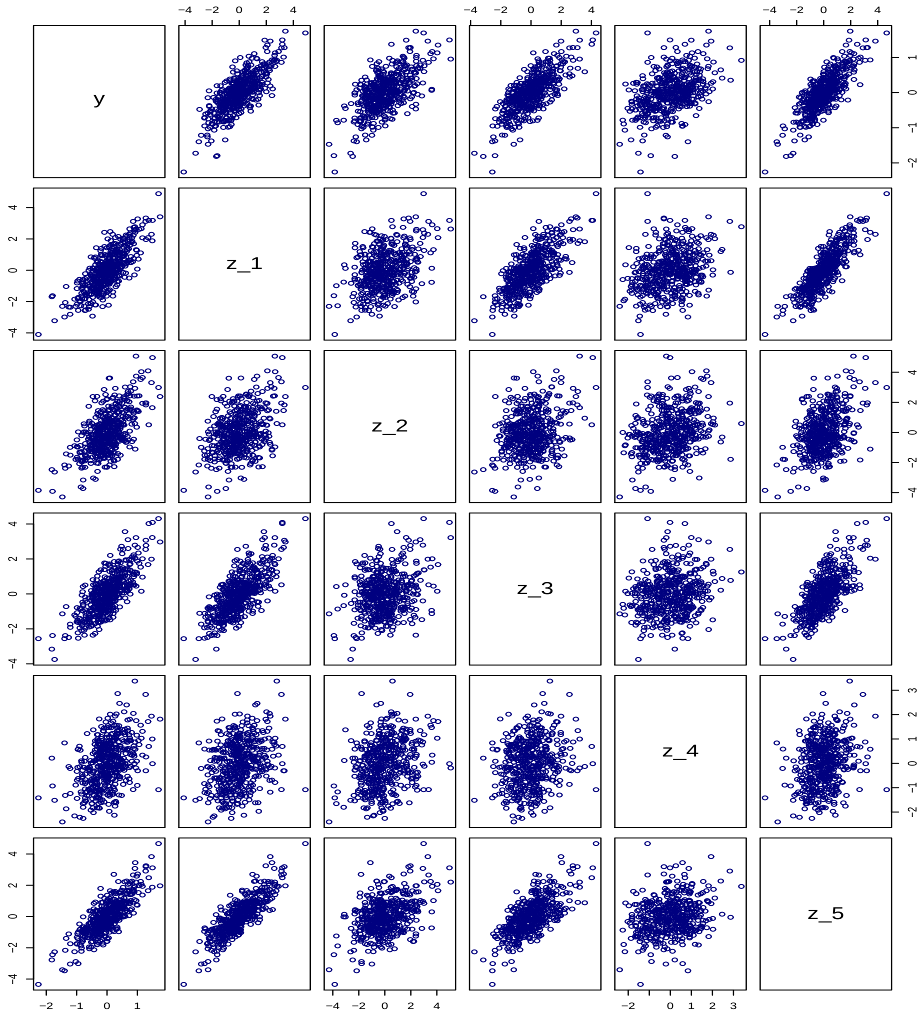


Figure 4. *The 25 scatterplots of the detrended true values of the observed predictand, y_t, and true values of the observed predictors, $\boldsymbol{z}_t$.*

6.1.2 Measurement errors

Measurement errors were simulated in *R* under assumptions that are supposed to mimic (with certain simplifications, see the discussion in [9], p.320) the characteristics of measurement errors in real-world data, namely errors in climate proxy data and instrumental observations. In this analysis all predictors represent proxies and instrumental data is represented by the predictand. The assumptions are:

- $\epsilon_t \sim$NI [2]$(0, \sigma_{\epsilon\epsilon})$
  Note, errors in instrumental data are not only uncorrelated, but also normally distributed, which implies the independence among errors.

[2] NI is an abbreviation for "distributed normally and independently"

- $\{u_{t,i}\}$ is an AR(1) process with zero mean and $\text{Var}(u_{t,i}) = \sigma_{u_i u_i}$ for all t and $i = 1, 2, 3, 4, 5$. When simulating AR(1) processes it was used that

  $$u_{1,i} = \frac{\delta_{1,i}}{\sqrt{(1-\rho^2)}},$$

  $$u_{t,i} = \rho \cdot u_{t-1,i} + \delta_{t,i}, \quad t > 1$$

  where $\delta_{t,i} \sim \text{NI}\Big(0, \sigma_{u_i u_i}(1-\rho^2)\Big)$ for all t and ρ is put to 0.5 or 0.8. Assuming $\delta_{t,i}$ normally distributed the uncorrelated assumption (see 3.2.1) automatically becomes an independence assumption.

- $\text{Cov}(\epsilon_t, u_{t,i}) = \text{Cov}(u_{t,i}, u_{t,j}) = 0, \quad \{i, j\} = 1, 2, \dots, 5, i \neq j$

- $\text{Cov}(\epsilon_t, u_{t-1,i})$ $=$Cov$(u_{t,i}, u_{t-1,j}) = 0$ for $t = 2, \dots, n$.

To implement simulations it is necessary to determine (theoretical) error variances, $\sigma_{\epsilon\epsilon}$, $\sigma_{u_i u_i}$, $i = 1, 2, 3, 4, 5$. Since each error term is supposed to account for a certain amount of the total variation in the corresponding observed variable, the variances were determined by the following relationship

$$\text{Var(error)} = \frac{\text{PNV}}{1 - \text{PNV}} \widehat{\text{Var}}(\text{corresponding sample of true values}),$$

where PNV (the Percent Noise by Variance, [12]) is defined as

$$\text{PNV} = \frac{\text{Var(error)}}{\widehat{\text{Var}}(\text{true values}) + \text{Var(error)}}$$

and its value is chosen in advance. The values of PNV used in the analysis are: 0.02, 0.1 for Y, which corresponds to 2% and 10% of the total variation in Y and 0.2, 0.5, 0.8 and 0.94 for Z, which corresponds to 20%, 50%, 80% and 94% of the total variation in Z_i, $i = 1, 2, 3, 4, 5$.

In Figure 5 one can see a certain realization of observed data, whose error terms satisfy the following combination: $\text{PNV}_Y = 0.02$, $\text{PNV}_Z = 0.5$, i.e. 50% of the total variation in each Z, and $\rho = 0.5$. Since the detrended true time series as well as the added error terms are stationary processes with constant variances, the observed time series are also stationary processes. Having in mind that one subset of observed data will be used for model fitting and the other for the prediction of Y, the fact that the observations follow the same model under the whole period appears to be an advantage.

Observed data has been divided in the following way. The first dataset covers the period from June 1861 to June 1990 (130 observations). That corresponds to a calibration period. The observations from this period will be denoted by the superscript (c). The second dataset covers the period from June 1501 to June 1860 (360 observations), which corresponds to a reconstruction period. The observations from this period will be denoted by the superscript (r). It is assumed that these two periods are independent, which may be in fact a simplification of reality.

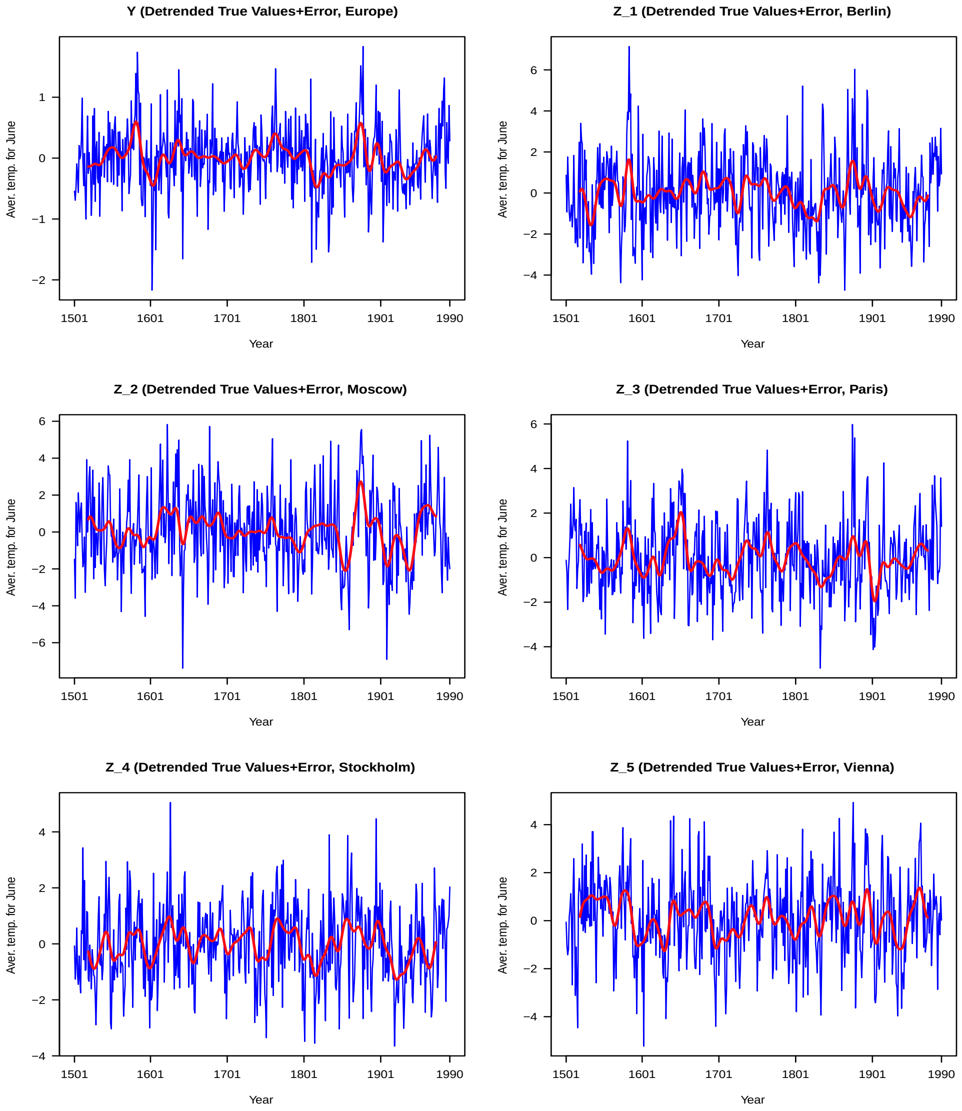


Figure 5. *The unsmoothed observed data in blue and the smoothed with gaussian filter observed data in red* ($\sigma_f = 9$).
$\mathrm{PNV}_Y = 0.02$, $\mathrm{PNV}_Z = 0.5$ *for each* Z *and* $\rho = 0.5$.

6.2 Assumptions and models

6.2.1 Assumption I - autocorrelated errors in predictors

Under the first (correct) assumption, measurement errors in the predictors are highly correlated - each sequance $\{u_{t,i}\}$, $i = 1, 2, 3, 4, 5$, constitutes an AR(1) process (for how it was simulated see section 6.1.2). However, by transforming the observed series $\{\boldsymbol{Z}\}$ in the following way

$$\begin{aligned}\boldsymbol{V}_1 = \sqrt{(1-\rho^2)}\boldsymbol{Z}_1 &= \underbrace{\sqrt{(1-\rho^2)}\boldsymbol{z}_1}_{=\boldsymbol{v}_1} + \underbrace{\sqrt{(1-\rho^2)}\boldsymbol{u}_1}_{=\boldsymbol{\delta}_1} \\ &= \boldsymbol{v}_1 + \boldsymbol{\delta}_1\end{aligned}$$

$$\begin{aligned}\boldsymbol{V}_t = \boldsymbol{Z}_t - \rho\boldsymbol{Z}_{t-1} &= \underbrace{\boldsymbol{z}_t - \rho\boldsymbol{z}_{t-1}}_{=\boldsymbol{v}_t} + \underbrace{\boldsymbol{u}_t - \rho\boldsymbol{u}_{t-1}}_{=\boldsymbol{\delta}_t} \\ &= \boldsymbol{v}_t + \boldsymbol{\delta}_t, \quad t > 1,\end{aligned}$$

for $\rho = \{0.5, 0.8\}$, the desired independence of the errors will be achived. Assuming that the true values y_t and $\boldsymbol{v}_t$ are perfectly related, it makes it possible to apply the measurement error regression with no error in the equation to the transformed data:

$$\begin{cases} y_t^{(c)} = \alpha_0 + \alpha_1 v_{t,1}^{(c)} + \alpha_2 v_{t,2}^{(c)} + \alpha_3 v_{t,3}^{(c)} + \alpha_4 v_{t,4}^{(c)} + \alpha_5 v_{t,5}^{(c)} \\ (Y_t^{(c)}, 1, \boldsymbol{V}_t^{(c)}) = (y_t^{(c)}, 1, \boldsymbol{v}_t^{(c)}) + (\epsilon_t^{(c)}, 0, \boldsymbol{\delta}_t^{(c)}) \\ Y_t^{(c)} = \boldsymbol{v'}_t^{(c)}\boldsymbol{\alpha} + \epsilon_t^{(c)} \end{cases}$$

where the vector of transformed true values $\boldsymbol{v}_t^{(c)}$ is treated as fixed in repeated sampling and $\boldsymbol{a} = (\epsilon_t^{(c)}, 0, \boldsymbol{\delta}_t^{(c)})'$ is the vector of measurement errors such that $\boldsymbol{a}_t \sim \mathrm{NI}(\boldsymbol{0}, \Sigma_{\boldsymbol{aa}})$, where

$$\begin{aligned}\Sigma_{\boldsymbol{aa}} &= \mathrm{diag}\Big(\sigma_{\epsilon\epsilon}, 0, \sigma_{\delta_1\delta_1}, \ldots, \sigma_{\delta_5\delta_5}\Big) \\ &= \mathrm{diag}\Big(\sigma_{\epsilon\epsilon}, 0, \sigma_{u_1u_1}(1-\rho^2), \ldots, \sigma_{u_5u_5}(1-\rho^2)\Big).\end{aligned}$$

Setting $V_{t,0} \equiv 1$ and $\mathrm{Var}(V_{t,0}) = \mathrm{Var}(\delta_{t,0}) = \sigma_{\delta_0\delta_0} = 0$, $\Sigma_{\boldsymbol{\delta\delta}}$ is the lower right 6 × 6 portion of $\Sigma_{\boldsymbol{aa}}$.

The choice of the functional model instead of the structural ME model, treating true values of predictors as random variables, seems most reasonable from a palaeoclimate perspective. Indeed, the required conditions for the structural model are normality and independency of the true values ([6], p.105). For the real-world time series, representing either instrumental measurements or proxies, the later condition is apparently inadequate. It worth mentioning that the described transformation warrants the independency of the transformed errors, not of the given transformed true time series.

In general, fitting multiple regression one has to be aware of the problem of (multi)collinearities among predictors, say $\boldsymbol{X}$. Collinearity arises if some predictors are exact or approximate linear combination of other, which leads to a singular or close to singular second-order matrix of predictors, $\boldsymbol{X}'\boldsymbol{X}$. This causes problems with the estimation of regression coefficients and affects their asymptotic properties. One way to detect collinearity is to examine the eigenvalues, $\boldsymbol{\lambda}$, of the matrix $\boldsymbol{X}'\boldsymbol{X}$. Small eigenvalues indicate a problem. Further one can calculate a condition number defined as

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}}.$$

It serves as a measure of the dispersion of the eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$. In general, if κ is less than 100, there are no serious problems with collinearity. Condition number between 100 and 1000 indicates from moderate to strong collinearity, while condition number larger than 1000 indicates serious collinearity ([10], p.301).

In case of multiple measurement error regression collinearity is defined in terms of the true values ([2]) and it affects first of all the asymptotic variance of the estimator of $\boldsymbol{\alpha}$, which involves the inverse of the second-order moment matrix of the true values, $\boldsymbol{M_{vv}} = \frac{1}{n}\sum_{t=1}^{130}(1, \boldsymbol{v}_t^{(c)})'(1, \boldsymbol{v}_t^{(c)})$ (see 3.1.4). The existence of collinearities among the unobservable true predictors will cause inflation of the variances. In practice one will wish to investigate the hypothesis that this matrix is positive definite. Under this analysis the true values are given, which allows us to examine the eigensystem of $\boldsymbol{M_{vv}}$ right away. The result has shown that when $\rho = 0.5$ and $\rho = 0.8$ the condition numbers of corresponding matrices are 17.4 and 16.5 respectively, that is under the critical value of 100. So the low degree of collinearity permits us to use all the transformed true variates $\boldsymbol{v}$ as true predictors.

Assuming that the true predictand and the true transformed predictors are not perfectly related, functional measurement error regression with an error in the equation will be applied to the transformed data:

$$\begin{cases} y_t^{(c)} = \alpha_0 + \alpha_1 v_{t,1}^{(c)} + \alpha_2 v_{t,2}^{(c)} + \alpha_3 v_{t,3}^{(c)} + \alpha_4 v_{t,4}^{(c)} + \alpha_5 v_{t,5}^{(c)} + q_t \\ (Y_t^{(c)}, 1, \boldsymbol{V}_t^{(c)}) = (y_t^{(c)}, 1, \boldsymbol{v}_t^{(c)}) + (\epsilon_t^{(c)}, 0, \boldsymbol{\delta}_t^{(c)}) \\ Y_t^{(c)} = \boldsymbol{v}_t'\boldsymbol{\alpha} + e_t^{(c)} \end{cases}$$

where $e_t^{(c)} = \epsilon_t^{(c)} + q_t^{(c)}$ and unknown random variable $q_t^{(c)}$ is NI(0,σ_{qq}).

6.2.2 Assumption II - uncorrelated errors in each predictor

Under this (incorrect) assumption no transformation of observed data is needed as under assumption I. Assuming that the true values y_t and $\boldsymbol{z}_t$ are perfectly related, Measurement Error (ME) regression with no error in the equation is applied to original data:

$$\begin{cases} y_t^{(c)} = \alpha_0 + \alpha_1 z_{t,1}^{(c)} + \alpha_2 z_{t,2}^{(c)} + \alpha_3 z_{t,3}^{(c)} + \alpha_4 z_{t,4}^{(c)} + \alpha_5 z_{t,5}^{(c)} \\ (Y_t^{(c)}, 1, \boldsymbol{Z}_t^{(c)}) = (y_t^{(c)}, 1, \boldsymbol{z}_t^{(c)}) + (\epsilon_t^{(c)}, 0, \boldsymbol{u}_t^{(c)}) \\ Y_t^{(c)} = \boldsymbol{z}'_t \boldsymbol{\alpha} + \epsilon_t^{(c)} \end{cases}$$

where the vector of known true values $\boldsymbol{z}_t^{(c)}$ is treated as fixed in repeated sampling and $\boldsymbol{a} = (\epsilon_t, 0, \boldsymbol{u}_t^{(c)})'$ is the vector of known measurement errors, which are independent of the true values $\boldsymbol{z}_t$, such that $\boldsymbol{a}_t \sim \mathrm{NI}(\boldsymbol{0}, \Sigma_{\boldsymbol{aa}})$, where

$$\Sigma_{\boldsymbol{aa}} = \mathrm{diag}\Big(\sigma_{\epsilon\epsilon}, 0, \sigma_{u_1 u_1}, \ldots, \sigma_{u_5 u_5}\Big).$$

Setting $Z_{t,0} \equiv 1$ and $\mathrm{Var}(Z_{t,0}) = \mathrm{Var}(u_{t,0}) = \sigma_{u_0 u_0} = 0$, $\Sigma_{\boldsymbol{uu}}$ becomes the lower right 6 × 6 portion of $\Sigma_{\boldsymbol{aa}}$.

The analysis of the eigensystem of $\boldsymbol{M}_{\boldsymbol{zz}} = \frac{1}{n}\sum_{t=1}^{130}(1, \boldsymbol{z}_t^{(c)})'(1, \boldsymbol{z}_t^{(c)})$ showed that its condition number is 22.7, which is much less than the critical value of 100. Hence, all five $\boldsymbol{z}_t^{(c)}$ can be included in the ME regression above.

Assuming that the true predictand and the true predictors are not perfectly related, the model under assumption II takes the form of functional measurement error regression with an error in the equation:

$$\begin{cases} y_t^{(c)} = \alpha_0 + \alpha_1 z_{t,1}^{(c)} + \alpha_2 z_{t,2}^{(c)} + \alpha_3 z_{t,3}^{(c)} + \alpha_4 z_{t,4}^{(c)} + \alpha_5 z_{t,5}^{(c)} + q_t \\ (Y_t^{(c)}, 1, \boldsymbol{Z}_t^{(c)}) = (y_t^{(c)}, 1, \boldsymbol{z}_t^{(c)}) + (\epsilon_t^{(c)}, 0, \boldsymbol{u}_t^{(c)}) \\ Y_t^{(c)} = \boldsymbol{z'}_t^{(c)} \boldsymbol{\alpha} + e_t^{(c)} \end{cases}$$

where $e_t = \epsilon_t + q_t$ and unknown random variable q_t is NI(0,σ_{qq}).

6.2.3 Assumption III - predictors are fixed

Assuming a linear relationship between observed values of Y_t and $\boldsymbol{Z}_t$ and the absence of measurement errors in the observed values of $\boldsymbol{Z}_t$, ordinary multiple linear regression is applied to original observed data:

$$Y_t^{(c)} = \alpha_0 + \alpha_1 Z_{t,1}^{(c)} + \alpha_2 Z_{t,2}^{(c)} + \alpha_3 Z_{t,3}^{(c)} + \alpha_4 Z_{t,4}^{(c)} + \alpha_5 Z_{t,5}^{(c)} + \epsilon_t^{(c)},$$

where $\epsilon_t^{(c)} \sim \mathrm{NI}(0, \sigma_{\epsilon\epsilon})$ are (known in this analysis) errors made in measuring y_t,

$\mathrm{E}[Y_t^{(c)}] = \alpha_0 + \alpha_1 Z_{t,1}^{(c)} + \alpha_2 Z_{t,2}^{(c)} + \alpha_3 Z_{t,3}^{(c)} + \alpha_4 Z_{t,4}^{(c)} + \alpha_5 Z_{t,5}^{(c)},$

$\mathrm{Var}(Y_t^{(c)}) = \sigma_{\epsilon\epsilon}$

$Z_{t,i}^{(c)}, i = 1.2. \ldots, 5$, are fixed numbers observed without measurement errors.

In contrast to ME regression collinearities under ordinary multiple regression is defined in terms of observed values and it affects both the estimation of the parameters and their variances. Knowing that the degree of collinearity among the given true variates is low, an even lower degree of collinearity among observed variates is expected. This expectation has actually been confirmed by inspection of condition numbers associated with 10000 matrices $\boldsymbol{Z'}^{(c)}\boldsymbol{Z}^{(c)}$ with $Z_0^{(c)} \equiv 1$, obtained for each combination of PNV_Z and ρ. As an example consider Figure 6, showing the distribution of condition numbers associated with 10000 matrices $\boldsymbol{Z'}^{(c)}\boldsymbol{Z}^{(c)}$ satisfying the combination: $\mathrm{PNV}_Z = 0.2$, $\rho = 0.5$. One can see that the maximal condition number is substantially less than the critical level of 100, which indicates a low degree of collinearity.

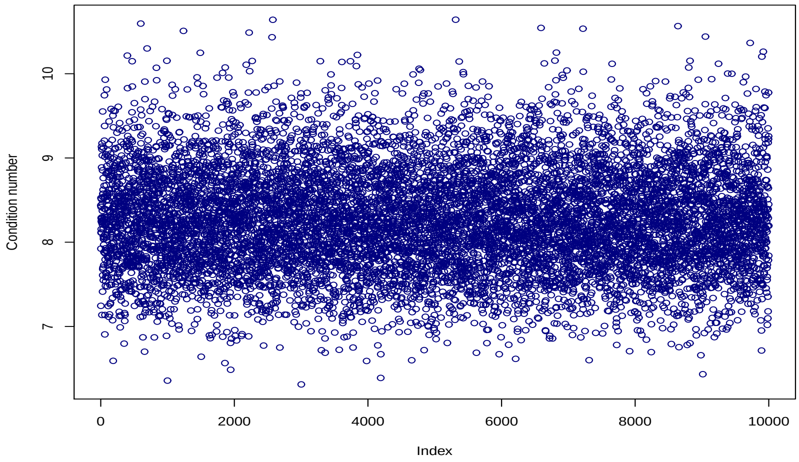


Figure 6. *Distribution of the largest condition number for 10000 matrices $\boldsymbol{Z'}^{(c)}\boldsymbol{Z}^{(c)}$, where each error term accounts for 20% of the total variation in corresponding Z_i, $i = 1, 2, \ldots, 5$ and $\rho = 0.5$.*

Regardless of the values of ρ, addition of measurement errors with larger variances has led to a much lower degree of collinearity than in the example above. Therefore, it can be concluded that all the five observed variables can be used simultaneously as predictors in the regression under assumption III.

6.3 Statistical analysis

In the previous section five models corresponding three assumptions about measurement errors in the predictors have been presented. The models are:

Model 1: ME regression `with no error` in the equation assuming autocorrelated measurement errors in predictors

Model 2: ME regression `with an error` in the equation assuming autocorrelated measurement errors in predictors

Model 3: ME regression `with no error` in the equation assuming uncorrelated measurement errors in each predictor

Model 4: ME regression `with an error` in the equation assuming uncorrelated measurement errors in each predictor

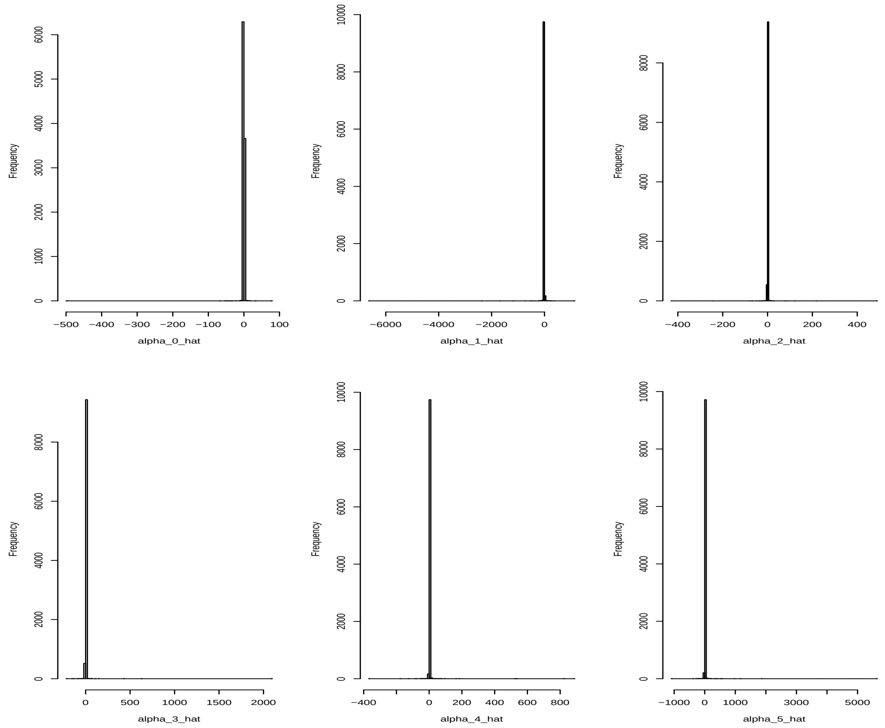
Model 5: Ordinary multiple regression assuming the absence of measurement errors in predictors.

This section is devoted to the analysis of the models. The idea is to fit them simultaneously to the same set of observed data with error terms simulated under assumptions described in section 6.1.2. Further by simulating error terms repeatedly, to form a large number of independent sets of observed data in order to use data from the calibration period from each of them for the simultaneous model fitting. This makes it possible to get an idea about the set of possible values of the estimators of $\boldsymbol{\alpha}$ under each model, given data with different magnitudes of the error variation. In the next stage of the analysis using data from the reconstruction period the predictive ability of the models will be assessed.

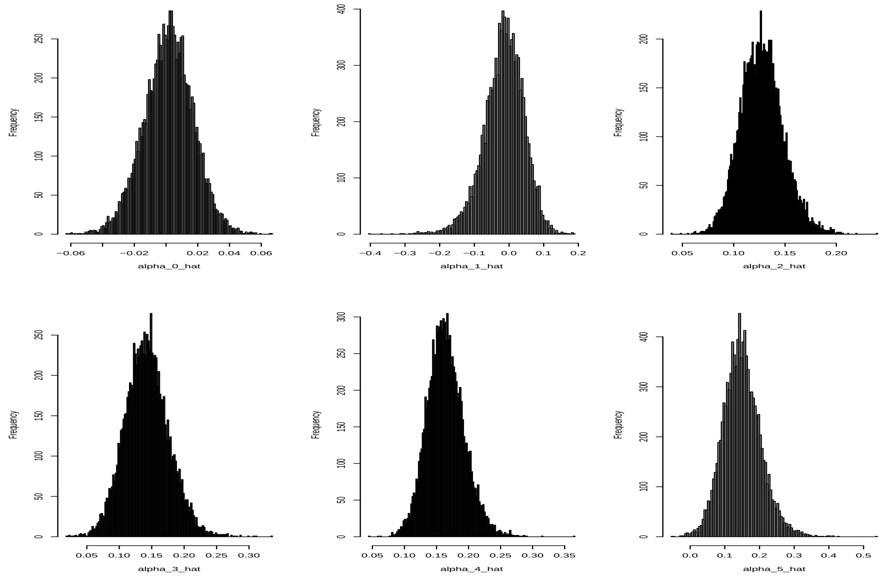
In total, 16 different types of observed data corresponding to 16 combinations of PNV$_Y$, PNV$_Z$ (for each Z simultaneously) and ρ, have been analysed. The combinations are:

	PNV$_Y$	ρ	PNV$_Z$		PNV$_Y$	ρ	PNV$_{\boldsymbol{Z}}$
1.	**0.02**	**0.5**	0.2	5.	**0.1**	**0.5**	0.2
2.			0.5	6.			0.5
3.			0.8	7.			0.8
4			0.94	8.			0.94
9.	**0.02**	**0.8**	0.2	13.	**0.1**	**0.8**	0.2
10.			0.5	14.			0.5
11.			0.8	15.			0.8
12.			0.94	16.			0.94

For each combination 10000 sets of observed data have been formed. Figures 7-11 contain histograms for 10000 maximum likelihood estimators for each model, based on data with PNV$_Y = 0.02$, $\rho = 0.5$ and PNV$_Z = 0.2$.



`Figure 7.` *Histogram for 10000 ML estimates of* $\boldsymbol{\alpha}$ *under model 1 when* $\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$, $\mathrm{PNV}_Z = 0.2$ *(for each Z).*



`Figure 8.` *Histogram for 10000 ML estimates of* $\boldsymbol{\alpha}$ *under model 2 when* $\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$, $\mathrm{PNV}_Z = 0.2$ *(for each Z).*

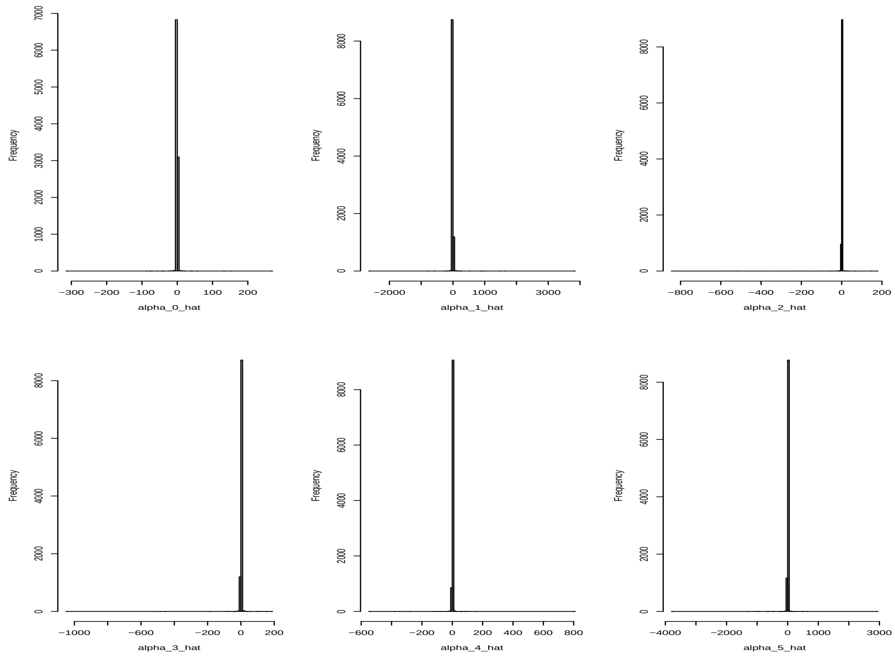


Figure 9. *Histogram for 10000 ML estimates of* $\boldsymbol{\alpha}$ *under model 3 when* $\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$, $\mathrm{PNV}_Z = 0.2$ *(for each Z).*

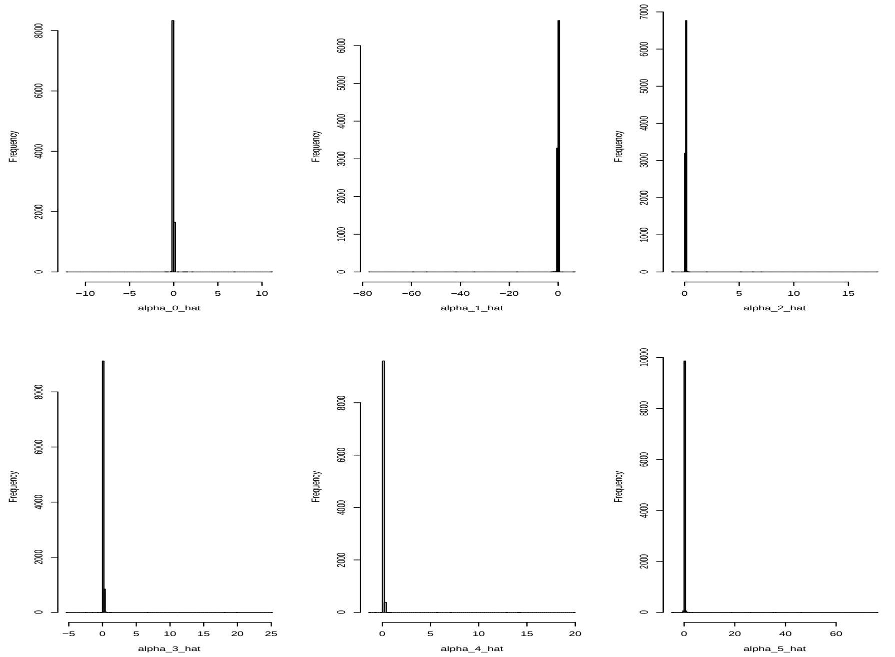
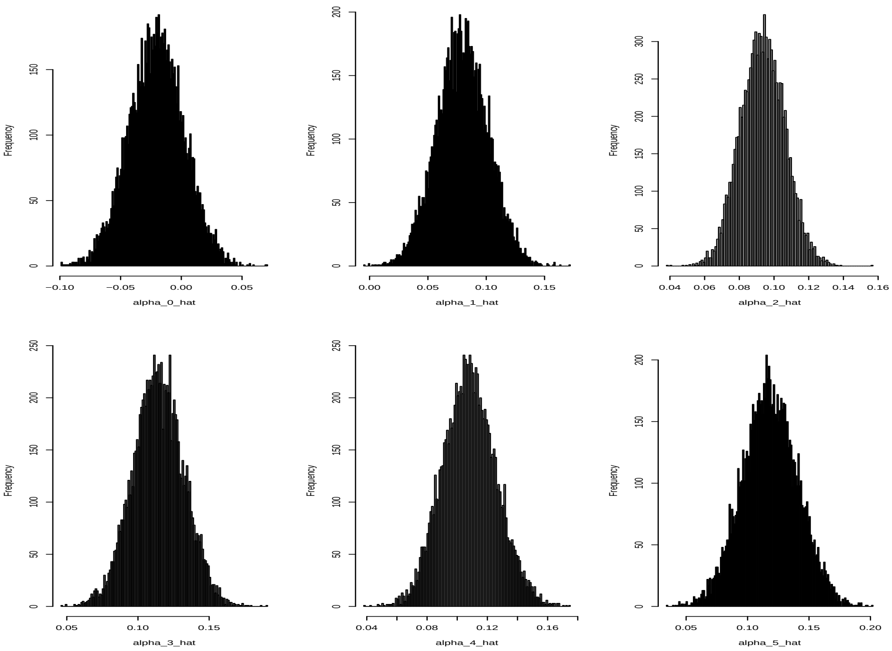


Figure 10. *Histogram for 10000 ML estimates of* $\boldsymbol{\alpha}$ *under model 4 when* $\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$, $\mathrm{PNV}_Z = 0.2$ *(for each Z).*



`Figure 11.` *Histogram for 10000 ML estimates of* $\boldsymbol{\alpha}$ *under model 5 when* $\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$, $\mathrm{PNV}_Z = 0.2$ *(for each Z).*

According to Figures 7-11 the qualitative difference between the marginal empirical distributions of the estimators under the five models is obvious. Under models 1, 3 and 4 extremly large values of the estimators of $\boldsymbol{\alpha}$ are observed. Of course, such values are unacceptable, in particular in the climate context. In contrast, under model 2 and 5 the obtained values seem to be quite reasonable. In the following table the results of the analogous analyses for data satisfying each of 16 combinations are summarized.

`Table 1.` The result of the simultaneous model fitting in terms of the obtained values of the estimators of $\boldsymbol{\alpha}$.

Model	Combinations for which (extremely) large estimators of $\boldsymbol{\alpha}$	
	have not been observed	have been observed
Model 1	-	for all combinations
Model 2	for $\mathrm{PNV}_Z = 0.2$ *(4 combinations)*	for $\mathrm{PNV}_Z \geq 0.5$ *(12 combinations)*
Model 3	-	for all combinations
Model 4	-	for all combinations
Model 5	for all combinations	-

It follows from Table 1 that among the two correct ME models (Model 1 and Model 2) it is only under Model 2 acceptable values of the estimators of $\boldsymbol{\alpha}$

have been observed in long run. Though only when the error variation is quite moderate, i.e. when it accounts for 20% of the total variation in each observed predictor ($\mathrm{PNV}_Z = 0.2$). As the error variation increases the problem of inconsistent estimation becomes actual even for this model. To understand the cause of this situation consider the estimators of $\boldsymbol{\alpha}$ under each model (see sections 5.1-2 and 6.2. for more details).

Model 1

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{M}_{\boldsymbol{V}^{(c)}\boldsymbol{V}^{(c)}} - \hat{\lambda}\Sigma_{\boldsymbol{\delta\delta}}\right)^{-1} \boldsymbol{M}_{\boldsymbol{V}^{(c)}Y^{(c)}}$$

$$= \left(\boldsymbol{M}_{\boldsymbol{V}^{(c)}\boldsymbol{V}^{(c)}} - \hat{\lambda}\Sigma_{\boldsymbol{uu}}(1-\rho^2)\right)^{-1} \boldsymbol{M}_{\boldsymbol{V}^{(c)}Y^{(c)}},$$

where $\hat{\lambda}$ is the smallest root of $\left|\boldsymbol{M}_{(Y^{(c)},\boldsymbol{V}^{(c)})(Y^{(c)},\boldsymbol{V}^{(c)})} - \lambda\boldsymbol{\Sigma}_{(\epsilon,\boldsymbol{\delta})(\epsilon,\boldsymbol{\delta})}\right| = 0,$

Model 2

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{M}_{\boldsymbol{V}^{(c)}\boldsymbol{V}^{(c)}} - \Sigma_{\boldsymbol{\delta\delta}}\right)^{-1} \boldsymbol{M}_{\boldsymbol{V}^{(c)}Y^{(c)}}$$

$$\left(\boldsymbol{M}_{\boldsymbol{V}^{(c)}\boldsymbol{V}^{(c)}} - \Sigma_{\boldsymbol{uu}}(1-\rho^2)\right)^{-1} \boldsymbol{M}_{\boldsymbol{V}^{(c)}Y^{(c)}},$$

Model 3

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{M}_{\boldsymbol{Z}^{(c)}\boldsymbol{Z}^{(c)}} - \hat{\lambda}\Sigma_{\boldsymbol{uu}}\right)^{-1} \boldsymbol{M}_{\boldsymbol{Z}^{(c)}Y^{(c)}},$$

where $\hat{\lambda}$ is the smallest root of $\left|\boldsymbol{M}_{(Y^{(c)},\boldsymbol{Z}^{(c)})(Y^{(c)}\boldsymbol{Z}^{(c)})} - \lambda\boldsymbol{\Sigma}_{(\epsilon,\boldsymbol{u})(\epsilon,\boldsymbol{u})}\right| = 0,$

Model 4

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{M}_{\boldsymbol{Z}^{(c)}\boldsymbol{Z}^{(c)}} - \Sigma_{\boldsymbol{uu}}\right)^{-1} \boldsymbol{M}_{\boldsymbol{Z}^{(c)}Y^{(c)}},$$

Model 5

$$\widehat{\boldsymbol{\alpha}} = \left(\boldsymbol{Z'}^{(c)}\boldsymbol{Z}^{(c)}\right)^{-1} \boldsymbol{Z'}^{(c)}Y^{(c)},$$

provided the matrices in the denominators are positive definite. The analysis of the eigensystem of 10000 corresponding matrices under each model and for each combination has shown that problems with estimation have a direct connection with violation of this condition. As a consequence the problems of negative $\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\alpha}})$, negative prediction errors and negative estimators of σ_{qq} under ME Error Equation model have arisen as well. In all cases when extremly large estimators were obtained the matrices were either close (sometimes extremely close) to singular or indefinite. For model 1 it was found out that unbounded denominators arose when $\hat{\lambda}$ was approximately equal to $\hat{\lambda}^*$, the smallest root of $\left|\boldsymbol{M}_{\boldsymbol{V}^{(c)}\boldsymbol{V}^{(c)}} - \lambda^*\boldsymbol{\Sigma}_{\boldsymbol{\delta\delta}}\right| = 0$, that is the smallest root of the determinantal equation for the denominator itself! The analogous situation was also typical for model 3.

As soon as the matrices in the denominators were positive definite and far away from being singular, bounded sets of obtained values of the estimators were observed as well as positive $\widehat{\text{Var}}(\widehat{\boldsymbol{\alpha}})$, positive prediction errors and positive estimators of σ_{qq} under ME Error Equation model. As mentioned earlier, it happened under Model 2, provided PNV$_Z$ = 0.2, and Model 5. Regarding model 5 this result is not surprising because 10000 matrices $\boldsymbol{Z'}^{(c)}\boldsymbol{Z}^{(c)}$ (for each combination) were inspected earlier in section 6.2.3.

In light of the obtained result, it seems reasonable to keep only two models for further comparative analysis. It is model 2 and 5. Their predictive abilities will be analysed by the aid of Mean Squared Error of Prediction, MSEP, provided data satisfy the four combinations of PNV values and ρ, namely

PNV$_Y$	ρ	PNV$_Z$
0.02	0.5	0.2
0.02	0.8	0.2
0.1	0.5	0.2
0.1	0.8	0.2

MSEP is a useful tool for assessing of model's prediction performance. The smaller the observed MSEP value is, the better predictive ability the model in question has. The definition of MSEP, based on separate validation data, is ([13], p.69)

$$MSEP = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2.$$

For the purpose of this analysis MSEP values were calculated on separate smoothed validation data denoted by the supercript (w)

$$\text{MSEP} = \frac{1}{360-4\sigma_f}\sum_{t=1}^{360-4\sigma_f}\left(\widehat{Y_t^{(r,w)}} - Y_t^{(r,w)}\right)^2, \quad 2\sigma_f + 1 \leq t \leq 360 - 2\sigma_f,$$

where

$$\widehat{Y_t^{(r,w)}} = \left\{\widehat{y_t^{(r,w)}}\right\} = \sum_{i=-2\sigma_f}^{2\sigma_f} w_i \widehat{Y_{t-i}^{(r)}} = \sum_{i=-2\sigma_f}^{2\sigma_f} w_i \boldsymbol{V'}_{t-i}^{(r)}\widehat{\boldsymbol{\alpha}} = \boldsymbol{V'}_t^{(r,w)}\widehat{\boldsymbol{\alpha}}$$

for model 2 and

$$\widehat{Y_t^{(r,w)}} = \left\{\widehat{y_t^{(r,w)}}\right\} = \sum_{i=-2\sigma_f}^{2\sigma_f} w_i \widehat{Y_{t-i}^{(r)}} = \sum_{i=-2\sigma_f}^{2\sigma_f} w_i \boldsymbol{Z'}_{t-i}^{(r)}\widehat{\boldsymbol{\alpha}} = \boldsymbol{Z'}_t^{(r,w)}\widehat{\boldsymbol{\alpha}}$$

for model 5. By letting the gaussian filters have only one term, $w = 1$ or equivalently $\sigma_f = 0$ in the expressions above, the data from the reconstruction period remain unsmoothed. In this analysis only smoothed data (the gaussian filter coefficient σ_f equal to 9) from the reconstruction period have been used. Note, that no smoothing was applied to the data from the calibration period when the parameters were estimated.

As the result of the analysis of MSEP values in long run, 10000 MSEP values have obtained under each model and for each combination. The range of the obtained values is given in the tabel below.

`Table 2.` *The range of 10000 MSEP values for model 2 and 5, given $PNV_Z = 0.2$.*

	$\text{PNV}_Y = 0.02, \rho = 0.5$		$\text{PNV}_Y = 0.02, \rho = 0.8$	
	Model 2	Model 5	Model 2	Model 5
min(MSEP)	0.01	0.0038	0.024	0.0041
max(MSEP)	0.033	0.0305	0.0393	0.0572

	$\text{PNV}_Y = 0.1, \rho = 0.5$		$\text{PNV}_Y = 0.1, \rho = 0.8$	
	Model 2	Model 5	Model 2	Model 5
min(MSEP)	0.01	0.005	0.023	0.006
max(MSEP)	0.0489	0.038	0.0483	0.072

The result presented in Table 2 suggests that there is no an essential difference in the MSEP ranges for the two models. No observed MSEP value gives a reason to reject any model as an unacceptable prediction method. Note, that smoothing in general increases the precision compared to unsmoothed data, implying larger MSEP values for smoothed data. However, in this case comparison with MSEP values calculated on unsmoothed data showed that the increase was modest.

Further, realizing that for each given true smoothed value from the reconstruction period, $y_t^{(r,w)}$, it can be constructed as many confidence intervals (CI's) as sets of observed data, i.e. 10000, appropriateness of the models can also be analysed by the aid of a long run analysis of CI's, in particular 90% CI's. The method used for construction of a 90% CI gives a CI which contains $y_t^{(r,w)}$ with probability 0.9 (for the formulas see Appendix), provided the distribution of $\widehat{\boldsymbol{\alpha}}$ is exact, as under Model 5. If the distribution of $\widehat{\boldsymbol{\alpha}}$ is asymptotic, as under Model 2, then the actual coverage probability can be either unknown in advance or unequal to 0.9. In addition the any use of an asymptotic confidence interval requires also knowledge of the expected length of the confidence interval. Under both models the coverage probability at each time t can be estimated according to the well-known recipe: the number of confidence intervals, containing $y_t^{(r,w)}$, is divided by the total number of confidence intervals available at time t. The expected length at time t can be estimated by averaging the lengths of all confidence intervals available at time t.

Althogether, 324 probabilities have been estimated (due to smoothing the number of the true values has decreased from 360 to 324). The result of estimations is the following. Under Model 2 depending on the combination between 68 and 72 estimated coverage probabilities were at least as large as 0.9. Many estimated probabilities were much lower than 0.9. This considerable variation is illustrated in Figure 13(a), given data satisfying $\text{PNV}_Y = 0.02$, $\text{PNV}_Z = 0.2$ (for each Z) and $\rho = 0.5$. Under Model 5 the highest estimated coverage probabilities for four combinations were not larger than 0.45. As an example consider Figure 13(b), showing the estimated coverage probabilities under Model 5 calculated on the same datasets as under Model 2 above.

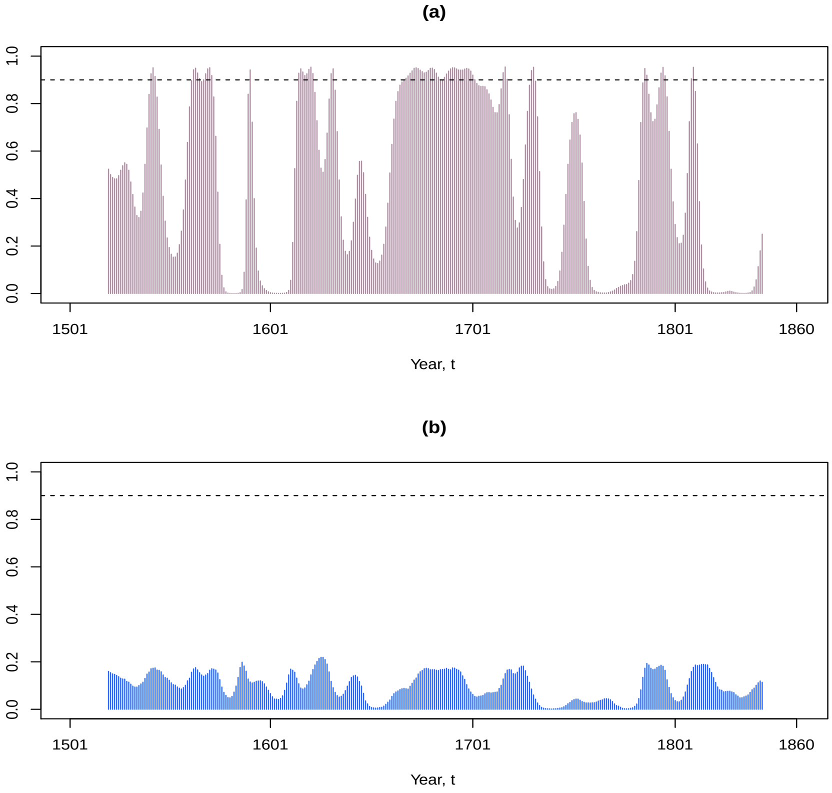


Figure 13. *Estimated coverage probabilities for each time t, $t \in \{Year : 1519, \ldots, 1842\}$ under model 2 in (a) and model 5 in (b) given smoothed data, ($\sigma_f = 9$), satisfying* $\mathrm{PNV}_Y = 0.02$, $\mathrm{PNV}_Z = 0.2$ *(for each Z)* and $\rho = 0.5$.
The dotted line marks the (nominal) conf. level 0.9.

Under Model 2 the expected length of a confidence interval with the nominal confidence level 0.9 have been studied as well. Figure 14 contains the plot of the mean length of 10000 confidence intervals calculated at each time t, given data satisfying $\mathrm{PNV}_Y = 0.02$, $\mathrm{PNV}_Z = 0.2$ (for each Z) and $\rho = 0.5$. Apparently, the variation in the mean length is much more stable compared to the variation in the estimated coverage probabilities under the same model (note, the scale at the y-axis). This means that two CI's with the nominal confidence level 0.9 and with almost equal lenghts can have very different coverage probabilities. The plot is dominated by a remarkable increase during the short time period $t = \{Year : 1812 - 1818\}$. The longest mean length, 0.1809, observed at time $t = \{Year : 1815\}$, corresponds to a quite low estimated coverage probability, namely 0.0233. The minimal mean length, 0.1617, observed at time $t = \{Year : 1710\}$ corresponds in its turn to a quite high estimated coverage probability, namely 0.7622. The similar variation pattern have also been observed for the remaining three combinations. In Table 3 the results about the range of the observed mean lengths, given data satisfying four combinations, are summarized.

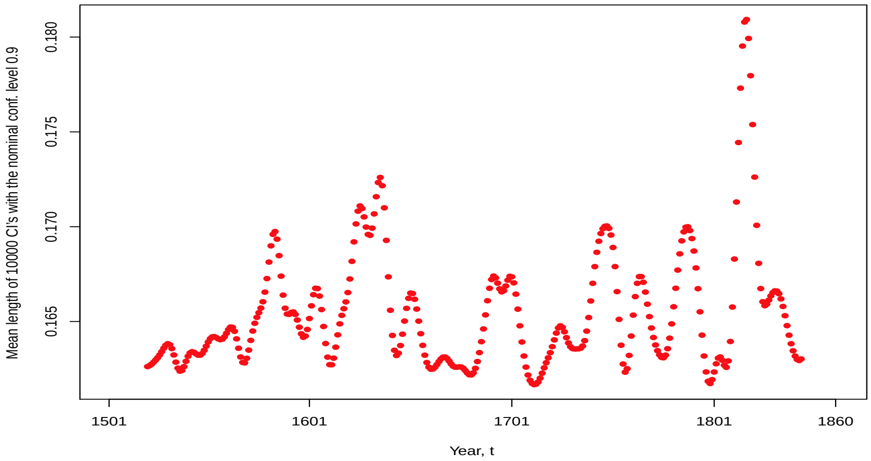


Figure 14. *Mean length of 10000 CI's with the nominal conf. level 0.9 at time t*, $t \in \{Year : 1519, \ldots, 1842\}$, *under Model 2, given smoothed data* $(\sigma_f = 9)$, *satisfying* $\mathrm{PNV}_Y = 0.02$, $\mathrm{PNV}_Z = 0.2$, $\rho = 0.5$.

Table 3. *The range of 324 mean lengths of 10000 CI's with the nominal conf. level 0.9 under Model 2 for 4 combinations with* $PNV_Z = 0.2$.

Combination	min(the mean length)	max(the mean length)
$\mathrm{PNV}_Y = 0.02$, $\rho = 0.5$	0.1617	0.1809
$\mathrm{PNV}_Y = 0.02$, $\rho = 0.8$	0.1101	0.1156
$\mathrm{PNV}_Y = 0.1$, $\rho = 0.5$	0.1733	0.1975
$\mathrm{PNV}_Y = 0.1$, $\rho = 0.8$	0.1221	0.1285

The results in Table 3 give some idea about expected length of a CI with the nominal confidence level 0.9 under Model 2, given data with the abovementioned properties. The short observed ranges, espessially for $\rho = 0.8$, allows one to determine the expected length of a CI with a quite high precision. The results in Table 3 suggests also that the expected length of a CI might be narrower as ρ increases and PNV_Y decreases.

It is clear that under Model 2, which takes into accout the variation in the predictors, a confidence interval with the nominal confidence level 0.9 at time t will be wider than a corresponding confidence interval under model 5. Nevertheless, two plots, illustrating this difference are presented (see Figure 15-16). All the calculations of the shown confidence intervals were carried out on the same set of observed data, given $\mathrm{PNV}_Y = 0.02$, $\mathrm{PNV}_Z = 0.2$, $\rho = 0.5$. Moreover, it was the first set of 10000 that satisfied the condition 0.9 <MSEP$_{Model2}$/MSEP$_{Model5}$ < 1.1, in other words, when the models' prediction performances are almost equal. Note that the known (observed) predictand, $Y_t^{(r,w)}$, is not plotted.

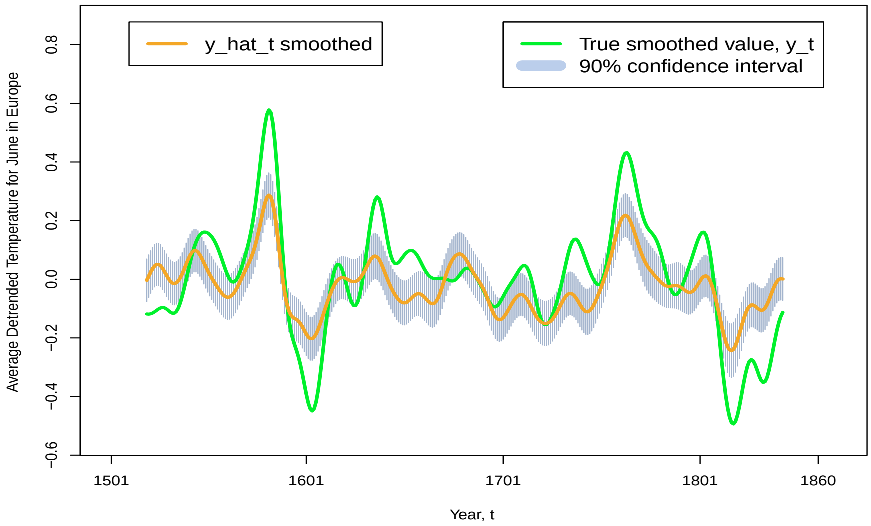


Figure 15. *Confidence interval with the nominal conf. level 0.9 for* $y_t^{(r,w)}$, $t \in \{Year : 1519, \ldots, 1842\}$ *under model 2 based on a certain set of observed data with* $\text{PNV}_Y = 0.02$, $\text{PNV}_Z = 0.2$ *and* $\rho = 0.5$. MSEP=0.01771.

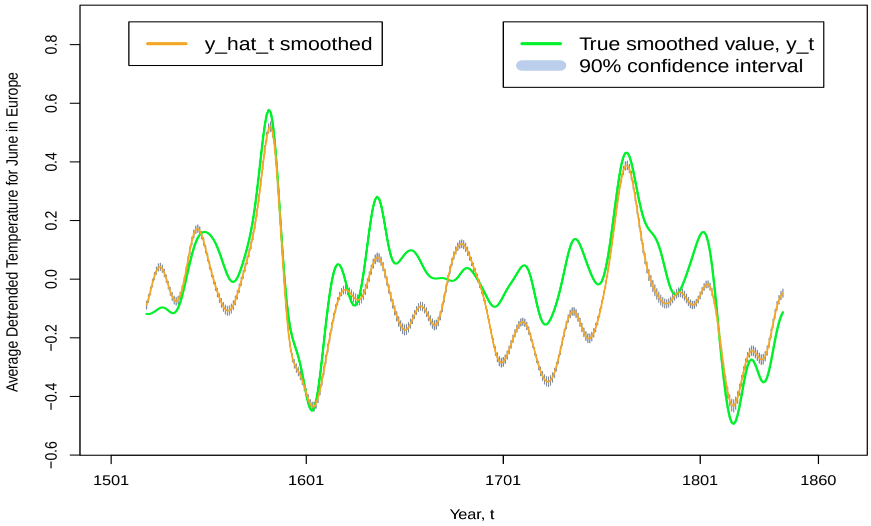


Figure 16. *90% confidence interval for* $y_t^{(r,w)}$, $t \in \{Year : 1519, \ldots, 1842\}$ *under model 5 based on the same set of observed data described in Figure 15.* MSEP=0.01736.

7 Discussion and Conclusions

This master thesis aims mainly to evaluate the performance of a multiple functional measurement error regression model with autocorrelated errors in predictors as a prediction method. The model assumes a linear relationship between true values of a predictand (temperature) and predictors (proxies). Furthermore, the model falls into two classes; model with an error in the equation and model with no error in the equation. The analysis is entirely based on synthetic data, formed in such a way that the assumptions of uncorrelated errors in the predictand and autocorrelated errors in the predictors were satisfied. In order to achieve independency of the observations, an appropriate transformation of the predictors has been performed. Based on Monte-Carlo methods, the analysis has indicated an inappropppriateness of ME model with no error in the equation (Model 1) as a prediction method regardless how large the error variation is and how strong the autocorrelation is. It seems that data do not support consistent estimation of the regression coefficients $\boldsymbol{\alpha}$. The main condition for consistent estimation - a positive definite matrix in the denominator in the estimators - was violated for the major part of the simulated sets of observed data. As a consequence large and extremely large estimators were observed.

The ME regression with an error in the equation (Model 2) seems to be adequate in terms of the obtained values of $\hat{\boldsymbol{\alpha}}$ though only when the error terms in the predictors account for 20% of the total variation in each observed predictor, i.e. $\text{PNV}_Z = 0.2$. Otherwise when measurement errors in the predictors account for at least 50% of the total variation in each observed predictor, the problem of inconsistent estimation becomes actual even for this model. Moreover, the number of such unacceptable estimators increases as the error variation increases, which points out that the existence of the matrix inverse becomes less likely as the error variation increases.

The analysis of ME regression with an error in the equation (Model 2), given that errors in the predictors account for 20% of the total variation in each observed predictor, has been continued in order to asses its prediction performance. It was done with help of Mean Squared Error of Prediction, MSEP. The observed minimum and maximum MSEP values (0.01 and 0.0489 respectively) gave no reason to reject this model as an unacceptable prediction method, although the conclusion to draw depends first of all on what criteria regarding MSEP values palaeoclimatologists have themselves. Further it can be noted that the variation in MSEP values exhibited a certain stability both when the autocorrelation ρ and the error variation in the predictand increases. The long run analysis of CI's with the nominal confidence level 0.9 at each time t in the reconstruction period has indicated, on the one hand, the rather narrow expected length of the desired CI, which appears to be an advantage, but on the other hand a great uncertainty about the actual coverage probability of the same CI. This uncertainty arises due to the large variation in observed possible values of the actual coverage probability. Only around 21% of the estimated coverage probabilities were at least as large as 0.9. All remaining estimated coverage probabilities turned out to be less or much less than the nominal confidence level 0.9. Because the smoothing of the data was taking into account when CI's were calculated, the same result is valid even for unsmoothed data. Hence, the observed results suggest to use this model as a prediction method with great caution, even if the error variation is modest and all assumptions about uncorrelated errors in istrumental data and autocorrelated errors in proxy data are satisfied.

In connection with the analysis of multiple (and univariate) ME models many questions regarding both the finite-sample properties of the estimator of $\boldsymbol{\alpha}$ such as unbiasedness, minimum variance and the existence/the absence of finite moments have arisen. But I have not found it possible to perform a detailed study of these questions within this project. It can be a topic of future projects, focusing only on multiple/univariate ME models.

The second aim of the thesis is to illustrate the inappropriateness of the use of models, which either do not take into consideration autocorrelation in measurement errors or do not allow for errors in predictors at all, when data in effect contain predictors with autocorrelated errors. These models were: ordinary multiple regression assuming fixed predictors and ME regression, both with no error in the equation and with an error in the equation, assuming uncorrelated errors in each predictor. They were fitted to the same datasets used for the analysis of the ME models with autocorrelated errors in predictors but without applying the transformation. The inappropriateness of the two ME models, assuming uncorrelated errors in each predictor, was detected at the first stage of the analysis. Large or extremely large values of the estimators of $\boldsymbol{\alpha}$ for the major part of available datasets were observed for all magnitudes of the error variation. Realizing that ME regression estimators obtained on ME Autocorrelated Errors model data are unreasonable estimators, this result is, in fact, sufficient to confirm the inappropriateness of this model as a prediction model when predictors are contaminated with autocorrelated errors.

In contrast, ordinary multiple regression estimators, which are both biased and inconsistent when they are obtained on any ME model data, has exhibited a good behaviour in terms of the observed values of $\hat{\boldsymbol{\alpha}}$ regardless how large the error variation is and how strong autocorrelation is. As a consequence only positive variances of the estimators and positive prediction errors have been observed. This depended on a linear independency between the observed predictors. So unless collinearity among observed predictors is present, this model can produce reasonable values of unreasonable estimators for the parameters of the ME model allowing for autocorrelated errors in predictors. Nevertheless, the inappropriateness of this model has been detected under the long run analysis of 90% confidence intervals. It turned out that no true single smoothed value of the predictand has the probability to be within a 90% confidence interval equal to at least 0.9. Moreover, the estimated probability for each of 324 true smoothed values was less than 0.45.

Under the analysis the problem of collinearities among true or observed predictors (depending on the model) has been taken into consideration. Thanks to the fact that only synthetic data with known properties were used, it was possible to detect with the firm confidence the absence of collinearities under each model. Regarding the ME model with autocorrelated errors it ensured that the variances of the estimators were not inflated, which implied reliable confidence intervals. On the whole, the use of synthetic data has substantially decreased the number of sources of uncertainty. In contrast to analysis based on real-world data, this analysis has not required estimation of autocorrelation ρ or testing the hypothesis if the second-order matrix of true values is positive definite. Further working with real-world data, the detrending of observed values does not necessary imply that true values will be detrended and will follow the same model during both the calibration period and the reconstruction period. In this analysis the availability of true values gave enough confidence in the assumption that the relationship between detrended true values of the predictand and detrended true values of the predictors is the same under both periods. All these factors together makes the results of this analysis highly reliable.

8 Appendix

8.1 Confidence interval under ordinary multiple linear regression

Given a vector of $k+1$ (smoothed) values of observed predictors from the reconstruction period, that is

$$\boldsymbol{X'}_t^{(r,w)} = \Big(1, X_{t,1}^{(r,w)}, X_{t,2}^{(r,w)}, \dots, X_{t,k}^{(r,w)}\Big),$$

a $100(1-p)\%$ confidence interval at the $(1-p)$ confidence level for the single true smoothed value $y_t^{(r,w)}$ is given by ([5])

$$\widehat{y_t^{(r,w)}} \pm t_{p/2}(n-k-1)\sqrt{\boldsymbol{X'}_t^{(r,w)}\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\alpha}})\boldsymbol{X}_t^{(r,w)}},$$

where

$$\widehat{\boldsymbol{\alpha}} = \Big(\boldsymbol{X'}^{(c)}\boldsymbol{X}^{(c)}\Big)^{-1}\boldsymbol{X'}^{(c)}Y^{(c)},$$

$$\widehat{\mathrm{Cov}}(\widehat{\boldsymbol{\alpha}}) = \hat{\sigma}_{\epsilon\epsilon}\Big(\boldsymbol{X'}^{(c)}\boldsymbol{X}^{(c)}\Big)^{-1}$$

and

$$\widehat{\boldsymbol{\alpha}} \sim \mathrm{N}\Big(\boldsymbol{\alpha}, \mathrm{Cov}(\widehat{\boldsymbol{\alpha}})\Big)$$

Note, that no smoothing is applied to the data from the calibration period. While when calculating the confidence interval the smoothed data from the reconstruction period can be used, that is

$$\boldsymbol{X'}_t^{(r,w)} = \sum_{i=-2\sigma_f}^{2\cdot\sigma_f} w_i \boldsymbol{X'}_{t-i}^{(r)}, \qquad 2\sigma_f + 1 \leq t \leq n - 2\sigma_f.$$

But letting the gaussian filters have only one term, $w = 1$, the data remain unsmoothed.

8.2 Confidence interval under multiple Measurement Error regression

Given a vector of observed (smoothed) values of $k+1$ predictors from the reconstruction period

$$\boldsymbol{X'}_t^{(r,w)} = \Big(1, X_{t,1}^{(r,w)}, X_{t,2}^{(r,w)}, \ldots, X_{t,k}^{(r,w)}\Big)',$$

a confidence interval with the nominal level $1-p$ for the single true smoothed value $y_t^{(r,w)}$ is given by

$$\widehat{y_t^{(r,w)}} \pm t_{p/2}(n-k-1))\sqrt{S_{\text{true}}^{(w)}},$$

where $S_{\text{true}}^{(w)}$ is an unbaised estimator of

$$\text{Var}\Big(\widehat{y_t^{(r,w)}} - y_t^{(r,w)}\Big) = \text{Var}\Big(\widehat{y_t^{(r,w)}}\Big) = \text{Var}\Big(\widehat{Y_t^{(r,w)}}\Big).$$

To obtained the expression of $\text{Var}\Big(\widehat{Y_t^{(r,w)}}\Big)$, multivariate forms of

$$\text{Var}(W_1) = \text{E}\Big[\text{Var}\Big(W_1|W_2\Big)\Big] + \text{Var}\Big(E\Big[W_1|W_2\Big]\Big)$$

and

$$E[W^2] = (\text{E}[W])^2 + \text{Var}(W),$$

have been used. It leads to

$$\begin{aligned}
\text{Var}\left(\widehat{Y_t^{(r,w)}}\right) &= \text{Var}\Big(\boldsymbol{X'}_t^{(r,w)}\widehat{\boldsymbol{\alpha}}\Big)\\
&= E\Big[\text{Cov}\Big(\boldsymbol{X'}^{(r,w)}\widehat{\boldsymbol{\alpha}}|\widehat{\boldsymbol{\alpha}}\Big)\Big] + \text{Cov}\Big(E\Big[\boldsymbol{X'}^{(r,w)}\widehat{\boldsymbol{\alpha}}|\widehat{\boldsymbol{\alpha}}\Big]\Big)\\
&= E\Big[\widehat{\boldsymbol{\alpha}}\boldsymbol{'}\text{Cov}\Big(\boldsymbol{X}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big)\widehat{\boldsymbol{\alpha}}\Big] + \text{Cov}\Big(E\Big[\boldsymbol{X'}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big]\widehat{\boldsymbol{\alpha}}\Big)\\
&= E[\widehat{\boldsymbol{\alpha}}'] \cdot \text{Cov}\Big(\boldsymbol{X}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big) \cdot E[\widehat{\boldsymbol{\alpha}}] + \text{tr}\left(\text{Cov}(\hat{\boldsymbol{\alpha}})\text{Cov}\Big(\boldsymbol{X}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big)\right) +\\
&\quad + E\Big[\boldsymbol{X'}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big] \cdot \text{Cov}(\widehat{\boldsymbol{\alpha}}) \cdot E\Big[\boldsymbol{X}_t^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big].
\end{aligned}$$

Estimating $E[\widehat{\boldsymbol{\alpha}}]$ and $E\Big[\boldsymbol{X}^{(r,w)}|\widehat{\boldsymbol{\alpha}}\Big]$ by $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{X}^{(r,w)}$, respectively, the estimated prediction error for the single true value becomes

$$\begin{aligned}
S_{\text{true}}^{(w)} =& \widehat{\boldsymbol{\alpha}}\boldsymbol{'}\text{Cov}\Big(\boldsymbol{X}_t^{(r,w)}\Big)\widehat{\boldsymbol{\alpha}} + \boldsymbol{X'}_t^{(r,w)}\widehat{\text{Cov}}(\widehat{\boldsymbol{\alpha}})\boldsymbol{X}_t^{(r,w)}\\
&+ \text{trace}\left(\widehat{\text{Cov}}(\widehat{\boldsymbol{\alpha}})\text{Cov}\left(\boldsymbol{X}_t^{(r,w)}\right)\right).
\end{aligned}$$

Smoothing of the data influence on the variances of observed variables in the followig way:

$$\mathrm{Var}\Big(X_{t,i}^{(r,w)}\Big) = \mathrm{Var}\Big(\sum_{l=-2\sigma_f}^{2\sigma_f} w_l X_{t-l,i}^{(r)}\Big) = \mathrm{Var}\Big(X_{t,i}^{(r)}\Big)\sum_{l=-2\sigma_f}^{2\sigma_f} w_l^2$$

for $i = 1, \ldots, k$ and $2\sigma_f + 1 \leq t \leq n - 2\sigma_f$,

$$\mathrm{Cov}\Big(X_{t,i}^{(r,w)}, X_{t,j}^{(r,w)}\Big) = \mathrm{Cov}\Big(\sum_{l=-2\sigma_f}^{2\sigma_f} w_l X_{t-l,i}^{(r)}, \sum_{s=-2\sigma_f}^{2\sigma_f} w_s X_{t-l,j}^{(r)}\Big)$$

$$= \sum_{l=-2\sigma_f}^{2\sigma_f} w_l \cdot \sum_{s=-2\sigma_f}^{2\sigma_f} w_s \cdot \mathrm{Cov}\Big(X_{t,i}^{(r)}, X_{t,j}^{(r)}\Big)$$

$$= 1 \cdot \mathrm{Cov}\Big(X_{t,i}^{(r)}, X_{t,j}^{(r)}\Big)$$

for $\{i, j\} = 1, \ldots, k$, $i \neq j$ and $2\sigma_f + 1 \leq t \leq n - 2\sigma_f$.

9 References

References

[1] P.J. Brockwell, R.A. Davis. *Introduction to time series and forecasting.* 2nd edition, Springer, 2002.

[2] O.Carrillo-Gamboa, R.F.Gunst. *Measurement error model colliniarities.* Technical Report No. SMU/DS/TR/240, Department of Statistical Science, Southern Methodist University, Dallas.

[3] Chi-Lun Cheng, J.W. Van Ness. *Statistical Regression with measurement error.* Kendall's Library Of Statistics, 1999.

[4] J.J.Faraway. *Practical Regression and Anova using R.* `www.stat.lsa.umich.edu/~faraway/book`, 2002.

[5] *Formulas collection for the course Applied Statistical Analysis* . Stockholm university, 2009.

[6] W.A. Fuller. *Measurement Error Models.* Wiley, 2006.

[7] J.J. Gomez-Navarro et al. *A 500 years-long simulation of the recent past over Europe*, work in progress.

[8] B. W. Lindgren. *Statistical theory. Fourth edition.* Chapman & Hall/CRC, 2000., p.270.

[9] A.Moberg, G. Brattström. *Prediction intervals for climate reconstructions with autocorrelated noise–An analysis of ordinary least squares and measurement error methods*, Journal `Palaeogeography,Palaeoclimatology, Palaeoecology`, 308 (2011) 313-329.

[10] D.C. Montgomery, E.A. Peck. *Introduction to Linear Regression Analysis*, Wiley, 1982.

[11] B. Pfaff. *Analysis of Integrated and Cointegrated Time series with R*, Springer, 2008.

[12] J.E.Smerdon. *Climate models as a test bed for climate reconstruction methods: pseudoproxy experiments*, WIREs Clim Change 2011.doi: 10.1002/wcc.149.

[13] R.Sundberg. *Compendium in Applied Mathematical Statistics.* Stockholm university, 2009.