Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors

Hamza Alkhatib, Boris Kargoll and Jens-André Paffenholz

Motivation & Objectives

Robust estimation is important in many fields of application where the probability density function of the random deviations is expected to be heavy-tailed (e.g., as a consequence of multiple outliers). In a multivariate regression model, each observable is modeled as a random vector which is explained by a vector-valued (possibly non-linear) deterministic regression function and a vector of random deviations. Besides heavy tails, multivariateness and non-linearity, a further aspect that complicates parameter estimation consists in the frequently encountered autocorrelatedness of the random deviations.

The purpose of this contribution is to study a time series model which can generally be described as the additive combination of a multivariate, nonlinear deterministic model with multiple univariate, covariance-stationary autoregressive (AR) processes whose white noise components follow independent scaled t-distributions. To obtain the maximum likelihood (ML) estimator, we derive a generalized expectation maximization (GEM) algorithm, which takes the form of linearized, iteratively reweighted least squares. As the degree of freedom of the t-distribution is also estimated, we have a partially adaptive estimate.

The observation model and the GEM algorithm

We model the uncertainties of the measurement process by means of random deviations between the observables and the functional model:

$$\mathbf{Y}_t = \mathbf{h}_t(\xi) + \mathbf{E}_t \qquad (t = 1, \ldots, n).$$

We assume that each of the N components of the random deviations is subject to autocorrelations in the form of a covariance-stationary autoregressive (AR) model

$$E_{k,t} = \alpha_{k,1}E_{k,t-1} + \ldots + \alpha_{k,p_k}E_{k,t-p_k} + U_{k,t} \qquad (k = 1, \ldots, N; t = 1, \ldots, n),$$

in which the random variables are, for every $k = 1,\ldots,N$, independently and identically t-distributed according to

$$U_{k,t} \sim t_{\nu_k}(0, \sigma_k^2) \qquad (k = 1, \ldots, N; t = 1, \ldots, n).$$

Results of real data application

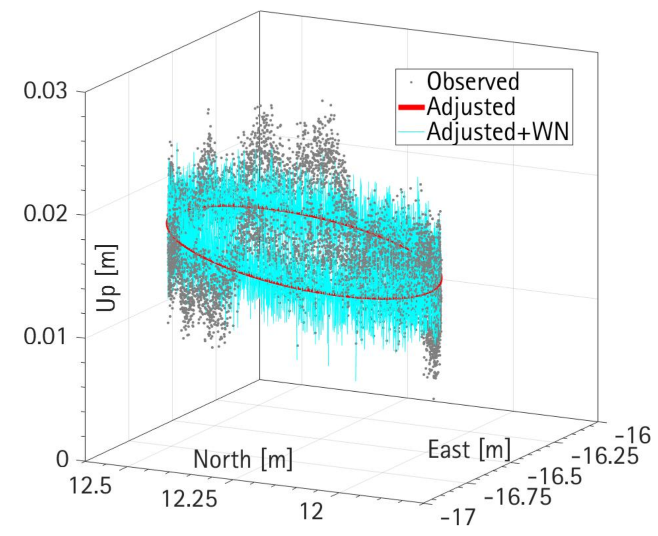
We extended the GEM algorithm based on [1] to approximate a measured and preprocessed 3D GNSS time series (see [2]) by a circle (see Alg. 1). One application of this model serves the geo-referencing of terrestrial laser scanner data where the 3D circle describes the circular, horizontal motion of two global navigation satellite system (GNSS) antenna reference points. Dual frequency receivers with individually and absolutely calibrated GNSS antennas were used (see Figure 1 top left). The origin of the coordinates lies in the nearby reference station with a baseline length of approximately 14 m. A full rotation consists of 7609 points (acquired with a data rate of 1 Hz) with respect to one antenna. We employed an AR model of order 15 for time series North- and East-component and of order 18 for time series Up-component. Figure 1 (top right) shows the adjusted circle and the observed 3D points. Having obtained an estimated degree of freedom of 10,000 for each component we conclude that given GNSS series are normally distributed.

Literature

[1] Kargoll, B., Omidalizarandi, M., Loth, I., Paffenholz, J.-A., Alkhatib, H.: An Iteratively Reweighted Least-Squares Approach to Adaptive Robust Adjustment of Parameters in Linear Regression Models with Autoregressive and t-Distributed Deviations. Journal of Geodesy, doi: 10.1007/s00190-017-1062-6, 2017.

[2] Paffenholz, J.-A.: Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. PhD thesis. Deutsche Geodätische Kommission (Series C, 689), Munich, http://www.dgk.badw.de/leadmin/docs/c-689.pdf, 2012.

Contact person: Dr.-Ing. Hamza Alkhatib, Geodetic Institute, Leibniz Universität Hannover, E-Mail: alkhatib@gih.uni-hannover.de



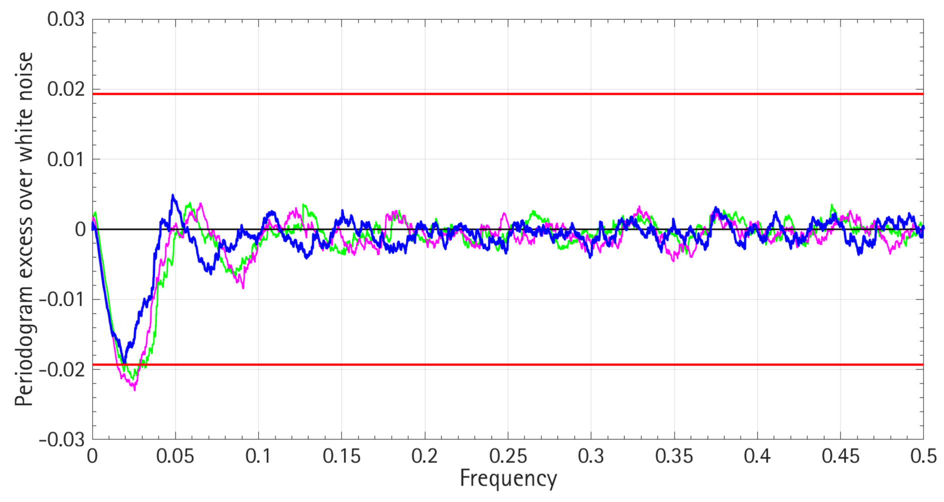


Figure 1: Upper left: Measurement setup. Upper right: 3D view of observed (grey points), adjusted circle (red line) and adjusted white noise (cyan) for n = 7827 real 3D-GNSS Observations; displayed in a North East Up (NEU) coordinate system. Bottom: Excess of the estimated periodogram of the decorrelated residuals for the North- and East- component AR(15) model (blue and green) and for the up-component AR(18) model (magenta) with respect to the theoretical white noise periodogram (black) and 99% significance bounds (red).

Algorithm 1: Generalized Expectation Maximization (GEM) algorithm

Input : multivariate time series $y_{k,t}$; functional models $h_{k,t}$; design matrices $\mathbf{A}_k$; Gauss-Newton step size $\gamma \in]0,1]$; with $k = 1, \ldots, N$ and $t = 1, \ldots, n$

Output: weights $w_{k,t}$;
parameter vector $\boldsymbol{\xi}$;
colored noise $e_{k,t}$;
AR coefficients $\boldsymbol{\alpha}_k$;
white noise $u_{k,t}$;
scale factors σ_k^2;
degrees of freedom $\nu_k^{(i+1)}$;

Initialization: $\boldsymbol{\xi}^{(0)}; \nu_k^{(0)}; \sigma_k^{(0)}; \boldsymbol{\alpha}_k^{(0)}$;

for $i = 1 \cdots \mathit{itermax}$ **do**

E-Step:

$$w_{k,t}^{(i)} = \frac{\nu_k^{(i)}+1}{\nu_k^{(i)} + \left(\frac{\alpha_k^{(i)}(L)(y_{k,t}-h_{k,t}(\xi^{(i)}))}{\sigma_k^{(i)}}\right)^2} \quad \rightsquigarrow \quad \mathbf{W}_k^{(i)} = \begin{bmatrix} w_{k,1}^{(i)} & 0 & \cdots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{k,n}^{(i)} \end{bmatrix}$$

M-Step:

$$\Delta y_{k,t}^{(i)} = y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i)}),\ \overline{\Delta y}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L)\Delta y_{k,t}, \text{ and } \overline{\mathbf{A}}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L)\mathbf{A}_{k,t}$$

$$\boldsymbol{\xi}^{(i+1)} = \boldsymbol{\xi}^{(i)} + \gamma \left(\sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\mathbf{A}}_k^{(i)} \right)^{-1} \sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\boldsymbol{\Delta}\mathbf{y}}_k^{(i)}$$

$$e_{k,t}^{(i+1)} = y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i+1)})$$

$$\mathbf{E}_k^{(i+1)} = \begin{bmatrix} e_{k,0}^{(i+1)} & \cdots & e_{k,1-p_k}^{(i+1)} \\ \vdots & & \vdots \\ e_{k,n-1}^{(i+1)} & \cdots & e_{k,n-p_k}^{(i+1)} \end{bmatrix}$$

$$\boldsymbol{\alpha}_k^{(i+1)} = \left((\mathbf{E}_k^{(i+1)})^T \mathbf{W}_k^{(i)} \mathbf{E}_k^{(i+1)} \right)^{-1} (\mathbf{E}_k^{(i+1)})^T \mathbf{W}_k^{(i)} \mathbf{e}_k^{(i+1)}$$

$$u_{k,t}^{(i+1)} = e_{k,t} - \alpha_{k,1}e_{k,t-1} - \ldots - \alpha_{k,p_k}e_{k,t-p_k} = \boldsymbol{\alpha}_k^{(i+1)}(L)e_{k,t}^{(i+1)}$$

$$(\sigma_k^2)^{(i+1)} = \frac{1}{n} \sum_{t=1}^{n} w_{k,t}^{(i)} \left(u_{k,t}^{(i+1)}\right)^2 = \frac{(\mathbf{u}_k^{(i+1)})^T \mathbf{W}_k^{(i)} \mathbf{u}_k^{(i+1)}}{n}$$

Solve: $\log \nu_k^{(i+1)} + 1 - \psi\left(\frac{\nu_k^{(i+1)}}{2}\right) + \psi\left(\frac{\nu_k^{(i+1)}+1}{2}\right) - \log\left(\nu_k^{(i+1)} + 1\right)$
$+\frac{1}{n} \sum_{t=1}^{n} \left(\log w_{k,t}^{(i+1)} - w_{k,t}^{(i+1)}\right) = 0$ with $\psi(.)$ digamma function