A Modified EM Algorithm for Parameter Estimation in Linear Models with Time-Dependent Autoregressive and t-Distributed Errors

Boris Kargoll[1], Mohammad Omidalizarandi[1], Hamza Alkhatib[1], and Wolf-Dieter Schuh[2]

[1] Leibniz Universität Hannover, Geodätisches Institut
Nienburger Str. 1, 30167 Hannover, Germany
{kargoll,zarandi,alkhatib}@gih.uni-hannover.de

[2] Rheinische Friedrich-Wilhelms-Universität Bonn,
Institut für Geodäsie und Geoinformation
Nussallee 17, 53115 Bonn, Germany
schuh@geod.uni-bonn.de

Abstract. We derive an expectation conditional maximization either (ECME) algorithm for estimating jointly the parameters of a linear regression model, of a time-variable autoregressive (AR) model with respect to the random deviations, and of a scaled t-distribution with respect to the white noise components. This algorithm is shown to take the form of iteratively reweighted least squares in the estimation of the parameters both of the regression and time-variability model. The fact that the degree of freedom of that distribution is also estimated turns the algorithm into a partially adaptive estimator. As low degrees of freedom correspond to heavy-tailed distributions, the estimator can be expected to be robust against outliers. It is shown that the initial stabilization phase of an accelerometer on a shaker table can be modeled parsimoniously and robustly by a Fourier series with AR errors for which the time-variability model is defined by cubic polynomials.

Keywords: Linear regression model, time-dependent AR process, partially adaptive estimation, robust parameter estimation, EM algorithm, iteratively reweighted least squares, scaled t-distribution

1 Introduction

Linear regression models are used in many fields of application to approximate numerical measurement results by means of parametric functions. In practice, the random deviations of the observables from these deterministic functions are frequently correlated. In the context of a time series measured by a single sensor, autocorrelations can be expected for instance as a consequence of calibration corrections being applied to all of the measurements (cf. [20]). The resulting colored noise is often modeled parametrically by means of covariance-stationary autoregressive (AR) or, more generally, by autoregressive moving average (ARMA) models. In the context of geodesy, for instance, such models were estimated to describe the colored noise of the Gravity and Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellite gravity gradiometer (see [31]), of inertial sensors (see [37], [29], [28]), and within global navigation satellite system (GNSS) data (cf. [23]). The idea of fusing a linear regression model with an AR error model has been known at least since [14].

Stationary colored noise models are frequently found to be insufficient due to time-variable effects acting on the sensor, the environment or the observed phenomenon. To overcome this limitation, AR(MA) processes with time-dependent coefficients were introduced and methods for their estimation investigated by [34] and [19]. If colored noise is to be removed from the measurements, the estimated AR(MA) process must be invertible. Invertibility conditions for time-variable ARMA processes were formulated by [12]. Many different schemes for modeling the time-variability of AR(MA) processes have been proposed. For instance, [16] used a stochastically perturbed difference equation constraint model to ensure smoothness of estimated time-variable AR processes. Another stochastic approach is based on the formulation of a time-variable AR process as a state-space model, leading to a Kalman filter (see [35] and the references therein). Furthermore, the modeling of AR models with coefficients changing throughout different regimes in the sense of a Markov chain was considered by [7]; see also the comparative study [1].

The usual approach to modeling the time-variability of AR(MA) coefficients is to assume a certain set of basis functions and to express a particular AR(MA) coefficient, at every time instance, as a point on the best-fitting linear combination of the basis functions. For instance, polynomials in terms of truncated power series [4], Legendre polynomials [30], wavelets [36], trigonometric [11, 6], sigmoid functions [10], and discrete prolate spheroidal sequences [8] have been used for this purpose. The previous studies made effective use of least squares techniques for the purpose of parameter estimation. When the white noise error component of the AR(MA) process is expected to be outlier-afflicted or heavy-tailed, a robust estimator should be used instead. Probability distributions that take care of both of these issues are found (amongst others) within the family of scaled t-distributions. When their degree of freedom is estimated alongside the regression or AR(MA) model parameters, one speaks then of partially adaptive estimation.

On the one hand, partially adaptive estimation for linear regression models with t-distributed random deviations was suggested by [18]. In their approach, an EM algorithm was used for the purpose of maximum likelihood (ML) estimation, which takes the form of numerically convenient iteratively reweighted least squares (as already indicated by [5]). [27] and [21] suggested expectation conditional maximization (ECM), expectation conditional maximization *either* (ECME) and multicycle ECM as variants of EM to speed up convergence. On the other hand, [3] carried out a Bayesian type of partially adaptive estimation of pure AR processes with t-distributed errors. [32] considered even a time-dependent AR process with t-distributed innovations, but did not estimate the degree of freedom. It should be mentioned that [25] and [24] introduced partially adaptive estimation, respectively, for linear regression and ARMA models in connection with a generalized t-distribution, which however does not seem to allow for the development of an EM algorithm in the spirit of [18].

Partially adaptive estimation of linear regression models with autoregressive random deviations and t-distributed white noise component appears to have been investigated first in [15]. In the following, this model is further extended to include an additional component that allows for time-variability of the AR coefficients. We choose for this purpose the aforementioned approach based on basis functions. After defining the specifics of the observation model, we derive a corresponding ECME algorithm, showing in particular that the coefficients of the regression model and the coefficients of the AR model can be estimated via two separate iteratively reweighted least squares schemes. This algorithm is applied to a measured time series of accelerometer data in a vibration analysis experiment. It is shown that the initial stabilization phase of the induced vibration can be modeled efficiently and estimated robustly by using a combination of a low-order Fourier series and a low-order, time-variable AR process with rather heavily tailed, t-distributed white noise components.

2 The Observation Model

The basic time series model consists of n linear observation equations

$$Y_t = \mathbf{A}_t \boldsymbol{\xi} + E_t \qquad (t = 1, \ldots, n), \tag{1}$$

where Y_t represents an observable, $\mathbf{A}_t \boldsymbol{\xi}$ a purely deterministic functional model, and E_t a random deviation. The observables and random deviations are collected in corresponding $(n \times 1)$-random vectors $\mathbf{Y}$ and $\mathbf{E}$. We assume that the functional model includes m unknown parameters $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_m]^T$ and that the row vectors $\mathbf{A}_1$, ..., $\mathbf{A}_n$ give rise to a known $(n \times m)$-coefficient matrix $\mathbf{A}$ with full rank m. Furthermore, we assume that the random deviations are autocorrelated via a time-dependent p-th order autoregressive (AR) model

$$E_t = \alpha_{1,t} E_{t-1} + \ldots + \alpha_{p,t} E_{t-p} + U_t \qquad (t = 1, \ldots, n), \tag{2}$$

where U_1, ..., U_n are independently and identically distributed random variables with mean 0 and variance σ_0^2. The time variability of each of the p AR coefficients will be described by linear models

$$\alpha_{j,t} = \mathbf{X}_t \boldsymbol{\beta}_j \qquad (j = 1, \ldots, p; t = 1, \ldots, n) \tag{3}$$

involving a $(q \times 1)$-vector of unknown parameters $\boldsymbol{\beta}_j = [\beta_{1,j}, \ldots, \beta_{q,j}]^T$ and known coefficient vectors $\mathbf{X}_1$, ..., $\mathbf{X}_n$. Thus, we employ the same family of functions for all p AR coefficients. We consider for the distribution of the white noise components $U_1, \ldots, U_n$ the scaled t-distribution $U_t \sim t_\nu(0, \sigma^2)$ with generally unknown degree of freedom ν and unknown scale parameter σ. That family of distributions is defined by the (family of) probability density functions (pdf)

$$f(u_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\sigma\,\Gamma\left(\frac{\nu}{2}\right)}\left[1 + \left(\frac{u_t}{\sigma}\right)^2/\nu\right]^{-\frac{\nu+1}{2}} \qquad (t = 1, \ldots, n), \tag{4}$$

where Γ is the gamma function. Due to the stochastic independence of the white noise components, their joint pdf factorizes into $f(\mathbf{u}) = \prod_{t=1}^{n} f(u_t)$. To fix the initial conditions of the AR(p) model (2), we assume that $E_0, \ldots, E_{1-p}$ take a constant value of 0. Assuming in addition that this model can be inverted, we can write in view of (1) – (3) for the t-th realization of U_t

$$\begin{aligned} u_t &= e_t - \alpha_{1,t}e_{t-1} - \ldots - \alpha_{p,t}e_{t-p} && (5)\\ &= (y_t - \mathbf{A}_t\boldsymbol{\xi}) - \mathbf{X}_t\boldsymbol{\beta}_1(y_{t-1} - \mathbf{A}_{t-1}\boldsymbol{\xi}) - \ldots - \mathbf{X}_t\boldsymbol{\beta}_p(y_{t-p} - \mathbf{A}_{t-p}\boldsymbol{\xi}) && (6)\\ &= \left(y_t - \sum_{j=1}^{m} A_{t,j}\xi_j\right) - \sum_{k=1}^{q} X_{t,k}\beta_{k,1}\left(y_{t-1} - \sum_{j=1}^{m} A_{t-1,j}\xi_j\right)\\ &\qquad - \ldots - \sum_{k=1}^{q} X_{t,k}\beta_{k,p}\left(y_{t-p} - \sum_{j=1}^{m} A_{t-p,j}\xi_j\right), && (7) \end{aligned}$$

setting the initial conditions $y_0 = \cdots = y_{1-p} = 0$ and $\mathbf{A}_0 = \cdots = \mathbf{A}_{1-p} = \mathbf{0}_{[1\times m]}$. Using the notation $L^j\mathbf{Z}_t := \mathbf{Z}_{t-j}$ in connection with $\boldsymbol{\alpha}_t(L) := 1 - \alpha_{1,t}L - \ldots - \alpha_{p,t}L^p$ and $\overline{\mathbf{Z}}_t = \alpha_t(L)\mathbf{Z}_t$ for an arbitrary sequence of matrices $(\mathbf{Z}_t)_{t\in T}$ with $T \subseteq \mathbb{Z}$, we also have for every $t = 1, \ldots, n$

$$u_t = \overline{e}_t = \boldsymbol{\alpha}_t(L)e_t = \boldsymbol{\alpha}_t(L)(y_t - \boldsymbol{A}_t\boldsymbol{\xi}) = \overline{y}_t - \overline{\mathbf{A}}_t\boldsymbol{\xi}. \tag{8}$$

This enables an interpretation of the quantities $\overline{e}_t$, $\overline{y}_t$ and $\overline{\mathbf{A}}_t$ as the outputs of the digital filter $\boldsymbol{\alpha}_t(L)$, applied respectively to a segment of the random deviations $\mathbf{e}$, of the observations $\mathbf{y}$ and of the coefficient matrix $\mathbf{A}$. Thus, $\boldsymbol{\alpha}_1(L), \ldots, \boldsymbol{\alpha}_n(L)$ may be viewed as turning the colored noise sequence $\mathbf{e}$ progressively into white noise $\mathbf{u}$, acting thus jointly as a *decorrelation filter*. Equations (4) and (7) define the basic probabilistic and parametric observation model. As $f(\mathbf{u})$ is actually a function of the observations $\mathbf{y}$, depending also on the values of all model parameters, we could use this joint pdf to define the likelihood function $\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2, \nu; \mathbf{y})$ for the purpose of parameter estimation. Note that this function is conditional on the previously fixed values for $E_0, \ldots, E_{1-p}$; the use of such a conditional likelihood function (cf. [13]) is justified if the number of observations is sufficiently large in order for the 'warm-up effect' of the initial conditions on the subsequent autoregressive values to fade out.

Since maximum likelihood (ML) estimation based on the preceding likelihood function (or on its natural logarithm) cannot be based on closed-form expressions due to the intricacy of the t-distribution, we apply a well-known latent-variables approach (see [18]), which will enable ML estimation by means of a relatively simple form of expectation maximization (EM) algorithm also for our specific time series model. The general idea is to firstly introduce independently and identically gamma-distributed latent variables $W_t \sim G(\nu/2, \nu/2)$ $(t = 1, \ldots, n)$, where ν is the degree of freedom of the desired t-distribution $t_\nu(0, \sigma^2)$. This distribution is defined by the pdf

$$f(w_t) = \begin{cases} \frac{(\frac{\nu}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \cdot w_t^{\frac{\nu}{2}-1} \cdot e^{-\frac{\nu}{2}w_t} & \text{if } w_t > 0, \\ 0 & \text{if } w_t \leq 0, \end{cases} \tag{9}$$

and the stochastic independence allows us to factorize $f(\mathbf{w}) = \prod_{t=1}^{n} f(w_t)$. Secondly, instead of assuming the white noise components to follow a scaled t-distribution at the outset, it is assumed that each random variable U_t follows a normal distribution conditional on the occurrence of the value w_t of the latent variable W_t. More specifically, we choose for the conditional pdf

$$f(u_t|w_t) = \frac{1}{\sqrt{2\pi(\sigma/\sqrt{w_t})^2}} \exp\left\{-\frac{u_t^2}{2(\sigma/\sqrt{w_t})^2}\right\}, \tag{10}$$

where each U_t is assumed to be conditionally independent from U_1, W_1, ..., U_{t-1}, W_{t-1}, U_{t+1}, W_{t+1}, ..., U_n and W_n. In other words, the values of the latter random variables shall not affect the density of u_t, in the sense that

$$f(u_t|u_1, w_1 \ldots, u_{t-1}, w_{t-1}, u_{t+1}, w_{t+1}, \ldots, u_n, w_n, w_t) = f(u_t|w_t). \tag{11}$$

This form of conditional independence can be interpreted as a hidden Markov property for which the hidden variables are real-valued (see Section 4 in [2]). In light of (10), we now see that the variance of every white noise component U_t is rescaled by an unknown (latent) weight w_t, independently of the white noise components and weights associated with time instances other than t. We obtain then from (9) and (10) the joint pdf $f(u_t, w_t) = f(w_t)\, f(u_t|w_t)$, which in turn gives the desired pdf (4) of the scaled t-distribution as a marginal distribution (see Sect. 2.6 in [26]). This joint pdf also yields $f(w_t, u_t) = f(u_t)\, f(w_t|u_t)$, where the conditional pdf $f(w_t|u_t)$ can be shown to define the gamma distribution $G(a, b)$ with parameters $a = (\nu + 1)/2$ and $b = (\nu + u_t^2/\sigma^2)/2$, given the value u_t (cf. [17], equations (27)). In connection with the initially made two assumptions of conditional independence, this allows us to establish the joint pdf of the white noise and the latent weights in the form of the factorization $f(\mathbf{u}, \mathbf{w}) = \prod_{t=1}^{n} f(w_t)\, f(u_t|w_t)$, which we define to be the likelihood function $\mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2, \nu; \mathbf{y}, \mathbf{w})$ of the extended observation model.

3 The Modified EM Algorithm

Combining the preceding pdf with (8) – (10), we can write the log-likelihood function in the form

$$\log \mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2, \nu; \mathbf{y}, \mathbf{w}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right)$$
$$-\frac{1}{2}\sum_{t=1}^{n}\log w_t - \frac{1}{2\sigma^2}\sum_{t=1}^{n} w_t[\boldsymbol{\alpha}_t(L)(y_t - \mathbf{A}_t\boldsymbol{\xi})]^2 + \frac{\nu}{2}\sum_{t=1}^{n}(\log w_t - w_t), \tag{12}$$

where each $\boldsymbol{\alpha}_t(L)$-filter is a function of the parameters $\boldsymbol{\beta}$. Collecting for brevity of expressions all unknown parameters $\boldsymbol{\xi}$, $\boldsymbol{\beta}$, σ^2 and ν within the vector $\boldsymbol{\theta}$ and following the idea of expectation maximization (EM) in [5], we iteratively aim for a solution $\boldsymbol{\theta}^{(i+1)}$ that maximizes $E_{\mathbf{Y},\mathbf{W}|\mathbf{y};\boldsymbol{\theta}^{(i)}}\left\{\log \mathcal{L}\left(\boldsymbol{\theta};\mathbf{y},\mathbf{W}\right)\right\}$. Here, $i \in \{0,1,2,\ldots,\}$ denotes the iteration step within the EM algorithm, so that the conditional expectation is evaluated by using both the given measurements $\mathbf{y}$ and the parameter estimates $\boldsymbol{\theta}^{(i)}$ from the preceding iteration step.

3.1 The E-Step

Following the general approach by [5], we restate the conditional expectation as the Q-function (see also [9])

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = E_{\mathbf{W}|\mathbf{y};\boldsymbol{\theta}^{(i)}}\left\{\log \mathcal{L}\left(\boldsymbol{\theta};\mathbf{y},\mathbf{W}\right)\right\}. \tag{13}$$

As the likelihood function was previously defined in terms of white noise $\mathbf{U}$ rather than the observables $\mathbf{Y}$, we will condition directly on the realizations $\mathbf{u}$. We then find with (12), in analogy both to the pure regression case without AR models in [18] and to the regression model with time-constant AR models in ([15]), the Q-function to be

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = &-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) \\ &-\sum_{t=1}^{n}\frac{1}{2}\left[\nu + \left(\frac{u_t}{\sigma}\right)^2\right]E_{\mathbf{W}|\mathbf{u};\theta^{(i)}}\{W_t\} + \sum_{t=1}^{n}\frac{1}{2}(\nu-1)E_{\mathbf{W}|\mathbf{u};\theta^{(i)}}\{\log W_t\} \end{aligned} \tag{14}$$

with

$$w_t^{(i)} := E_{\mathbf{W}|\mathbf{u};\theta^{(i)}}\{W_t\} = \frac{\nu^{(i)}+1}{\nu^{(i)} + \left(\frac{\boldsymbol{\alpha}_t^{(i)}(L)(y_t - \mathbf{A}_t\xi^{(i)})}{\sigma^{(i)}}\right)^2} \tag{15}$$

and (employing the digamma function ψ)

$$E_{\mathbf{W}|\mathbf{u};\theta^{(i)}}\{\log W_t\} = \log w_t^{(i)} + \psi\left(\frac{\nu^{(i)}+1}{2}\right) - \log\left(\frac{\nu^{(i)}+1}{2}\right). \tag{16}$$

Substitution of the findings (15) and (16) into (14) gives us now

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = &\text{const.} - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{n} w_t^{(i)}\left[\boldsymbol{\alpha}_t(L)(y_t - \mathbf{A}_t\boldsymbol{\xi})\right]^2 + \frac{n\nu}{2}\log\nu \\ &-n\log\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\left[\psi\left(\frac{\nu^{(i)}+1}{2}\right) - \log\left(\nu^{(i)}+1\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log w_t^{(i)} - w_t^{(i)}\right)\right]. \end{aligned} \tag{17}$$

3.2 The M-Step

To carry out the M-Step, we determine the first partial derivatives of the Q-function (17) with respect to the individual parameters $\boldsymbol{\xi}$, $\boldsymbol{\beta}$, σ^2 and ν in $\boldsymbol{\theta}$, and set these equal to zero. It is not difficult to show that the first-order condition with respect to the j-th parameter in $\boldsymbol{\xi}$ becomes

$$0 = \frac{\partial}{\partial \xi_j} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \frac{1}{\sigma^2} \sum_{t=1}^{n} w_t^{(i)} \overline{A}_{t,j} (\overline{y}_t - \overline{\mathbf{A}}_t \boldsymbol{\xi}).$$

Writing these m equations in matrix notation, for which purpose we denote by $\mathbf{W}^{(i)}$ the diagonal matrix of the weights $w_1^{(i)}$, …, $w_n^{(i)}$, we obtain

$$\mathbf{0} = \begin{bmatrix} \overline{A}_{1,1} & \cdots & \overline{A}_{n,1} \\ \vdots & & \vdots \\ \overline{A}_{1,m} & \cdots & \overline{A}_{n,m} \end{bmatrix} \mathbf{W}^{(i)} \begin{bmatrix} \overline{y}_1 - \overline{\mathbf{A}}_1 \boldsymbol{\xi} \\ \vdots \\ \overline{y}_n - \overline{\mathbf{A}}_n \boldsymbol{\xi} \end{bmatrix} = \overline{\mathbf{A}} \mathbf{W}^{(i)} \left(\overline{\mathbf{y}} - \overline{\mathbf{A}} \boldsymbol{\xi} \right).$$

As these normal equations for the parameter group $\boldsymbol{\xi}$ involve also the unknown parameters $\boldsymbol{\beta}$ through the filter operations, we fix values for the latter by setting $\boldsymbol{\beta} = \boldsymbol{\beta}^{(i)}$. In doing this, we perform a so-called conditional maximization (CM) step in the sense of [27]. Then, $\boldsymbol{\beta}^{(i)}$ allows us to compute the time variable AR coefficients $\alpha_{1,1}^{(i)}$, …, $\alpha_{p,n}^{(i)}$ by means of the equations (3); these coefficients define decorrelation filters, which we can subsequently employ to calculate the filtered quantities (for every $t = 1, \ldots, n$)

$$\overline{y}_t^{(i)} := \boldsymbol{\alpha}_t^{(i)}(L) y_t, \qquad \overline{A}_{t,j}^{(i)} := \boldsymbol{\alpha}_t^{(i)}(L) A_{t,j}, \qquad \overline{\mathbf{A}}_t^{(i)} := \boldsymbol{\alpha}_t^{(i)}(L) \mathbf{A}_t. \tag{18}$$

The new solution $\boldsymbol{\xi}^{(i+1)}$ for $\boldsymbol{\xi}$ can then be computed from

$$\boldsymbol{\xi}^{(i+1)} = \left((\overline{\mathbf{A}}^{(i)})^T \mathbf{W}^{(i)} \overline{\mathbf{A}}^{(i)} \right)^{-1} (\overline{\mathbf{A}}^{(i)})^T \mathbf{W}^{(i)} \overline{\mathbf{y}}^{(i)}, \tag{19}$$

which estimates give rise to the colored noise residuals $e_t^{(i+1)} := y_t - \mathbf{A}_t \boldsymbol{\xi}^{(i+1)}$. Next, we consider the first-order conditions with respect to the previously fixed parameter vectors $\boldsymbol{\beta}_1$, …, $\boldsymbol{\beta}_p$, for which we obtain

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{\beta}_h} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = \frac{1}{\sigma^2} \sum_{t=1}^{n} w_t^{(i)} e_{t-h} \mathbf{X}_t^T (e_t - \mathbf{X}_t \boldsymbol{\beta}_1 e_{t-1} - \ldots - \mathbf{X}_t \boldsymbol{\beta}_p e_{t-p}).$$

Having already determined estimates $\boldsymbol{\xi}^{(i+1)}$, the joint solution of the equation systems arising for all $h = 1, \ldots, p$ can be determined as a second CM step via

$$\mathbf{0} = \begin{bmatrix} e_0^{(i+1)} \mathbf{X}_1^T & \cdots & e_{n-1}^{(i+1)} \mathbf{X}_n^T \\ \vdots & & \vdots \\ e_{1-p}^{(i+1)} \mathbf{X}_1^T & \cdots & e_{n-p}^{(i+1)} \mathbf{X}_n^T \end{bmatrix} \mathbf{W}^{(i)} \begin{bmatrix} e_1^{(i+1)} - e_0^{(i+1)} \mathbf{X}_1 \boldsymbol{\beta}_1 - \ldots - e_{1-p}^{(i+1)} \mathbf{X}_1 \boldsymbol{\beta}_p \\ \vdots \\ e_n^{(i+1)} - e_{n-1}^{(i+1)} \mathbf{X}_n \boldsymbol{\beta}_1 - \ldots - e_{n-p}^{(i+1)} \mathbf{X}_n \boldsymbol{\beta}_p \end{bmatrix}$$
$$=: (\mathbf{E}^{(i+1)})^T \mathbf{W}^{(i)} \left(\mathbf{e}^{(i+1)} - \mathbf{E}^{(i+1)} \boldsymbol{\beta} \right),$$

using the initial conditions $e_0^{(i+1)} = \ldots = e_{1-p}^{(i+1)} = 0$ and the stacked vector $\boldsymbol{\beta}^T = [\boldsymbol{\beta}_1^T \ldots \boldsymbol{\beta}_p^T]$. The reweighted least squares solution for $\boldsymbol{\beta}$ then reads

$$\boldsymbol{\beta}^{(i+1)} = \left((\mathbf{E}^{(i+1)})^T \mathbf{W}^{(i)} \mathbf{E}^{(i+1)}\right)^{-1} (\mathbf{E}^{(i+1)})^T \mathbf{W}^{(i)} \mathbf{e}^{(i+1)}. \tag{20}$$

For every time instance t, the resulting AR coefficients will be denoted by $\alpha_{j,t}^{(i+1)}$, and we write for the corresponding decorrelation filter $\boldsymbol{\alpha}_t^{(i+1)}(L)$, which allows us to estimate the white noise residuals through $u_t^{(i+1)} = \boldsymbol{\alpha}_t^{(i+1)}(L)e_t^{(i+1)}$. The third CM-Step applies to the scale factor of the underlying t-distribution and requires the solution of

$$0 = \frac{\partial}{\partial \sigma^2} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{n} w_t^{(i)} u_t^2.$$

Substituting the current estimates $\boldsymbol{\xi}^{(i+1)}$ and $\boldsymbol{\beta}^{(i+1)}$, we obtain for this solution the average sum of squared residuals

$$(\sigma^2)^{(i+1)} = \frac{1}{n} \sum_{t=1}^{n} w_t^{(i)} \left(u_t^{(i+1)}\right)^2 = \frac{(\mathbf{u}^{(i+1)})^T \mathbf{W}^{(i)} \mathbf{u}^{(i+1)}}{n}. \tag{21}$$

The fourth CM step would then follow from solving the equation $0 = \frac{\partial}{\partial \nu} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)})$, completing the current step of the EM algorithm (in the form of ECM). In view of the findings of [21] and [22] (see also [26]) in the context of estimating the degree of freedom of the scaled t-distribution, the number of iteration steps can generally be reduced greatly by replacing the Q-function with the original log-likelihood function within the preceding first-order condition for ν. This modification of the ECM algorithm is called ECM *either* (ECME) and increases the likelihood in each iteration step as well. Applying this idea to our specific model (4) – (8), we thus seek the zero of the equation

$$\begin{aligned} 0 = \frac{\partial}{\partial \nu} \log \mathcal{L}(\boldsymbol{\xi}, \boldsymbol{\beta}, \sigma^2, \nu; \mathbf{y}) &= \frac{n}{2}\psi\left(\frac{\nu+1}{2}\right) - \frac{n}{2}\psi\left(\frac{\nu}{2}\right) + \frac{n}{2}(\log \nu + 1) \\ &- \frac{1}{2}\sum_{t=1}^{n} \log\left[\nu + \left(\frac{u_t}{\sigma}\right)^2\right] - \frac{1}{2}(\nu+1)\sum_{t=1}^{n}\left[\nu + \left(\frac{u_t}{\sigma}\right)^2\right]^{-1} \end{aligned}$$

Replacing $\mathbf{u}$ by the currently available estimated residuals $\mathbf{u}^{(i+1)}$, denoting the solution for ν by $\nu^{(i+1)}$, and defining $w_t^{(i+1)}$ according to (15), we finally obtain

$$\begin{aligned} 0 = \log \nu^{(i+1)} + 1 - \psi\left(\frac{\nu^{(i+1)}}{2}\right) + \psi\left(\frac{\nu^{(i+1)}+1}{2}\right) - \log\left(\nu^{(i+1)}+1\right) \\ + \frac{1}{n}\sum_{t=1}^{n}\left(\log w_t^{(i+1)} - w_t^{(i+1)}\right). \end{aligned} \tag{22}$$

We conclude this section with a few comments on our implementation of the preceding ECME algorithm.

If initial values for $\boldsymbol{\xi}$ are unknown, they are computed via unweighted least squares through $\boldsymbol{\xi}^{(0)} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. Based on the resulting initial residuals $\mathbf{e}^{(0)}$, initial values for $\boldsymbol{\beta}$ are computed next via (20), applying again the neutral weight matrix $\mathbf{W}^{(0)} = \mathbf{I}$. This solution gives rise to the initial decorrelation filters $\alpha_t^{(0)}(L)$, which allow for the computation of the residuals $\mathbf{u}^{(0)}$. Subsequently, we determine the initial value for σ^2 through $(\sigma^2)^{(0)} = (\mathbf{u}^{(0)})^T\mathbf{u}^{(0)}/n$. Furthermore, we choose for the initial degree of freedom $\nu^{(0)} = 30$. With these values, the weight determination (15) within the E step and the CM(E) steps (19) – (22) are iterated until the maximum number of iterations is reached or until the following stop criterion is satisfied. We check whether the greatest absolute value of the differences between the estimates of two subsequent iteration steps is less than 10^{-8} for the parameters $\boldsymbol{\xi}$, $\boldsymbol{\beta}$ and σ^2, and less than 10^{-4} for ν. Since the normal distribution represents the limiting case $\nu \to \infty$ of the t-distribution, it is possible that the zero of (22) is infinite. To circumvent numerical problems created by this case, we check if a sign change of the function on the right-hand side of (22) occurs between 10^{-8} and 10^8; if not, we set $\nu^{(i+1)}$ to a very large value in correspondence to a normal distribution. It should be mentioned that we did not find it necessary in our real-data applications to enforce stability on the time-variable AR-processes, which issue is beyond the scope of the current paper.

4 An Application to Vibration Analysis

We applied the ECME algorithm to estimate the non-stationary behavior of a highly accurate single-axis PCB Piezotronics accelerometer within a vibration analysis experiment, which was carried out at the Institute of Concrete Construction at the Leibniz Universität Hannover. The sensor was mounted on a shaker table, which consists of a plexiglass plate fixed between two wooden supports and two imbalance motors in the center. This shaker induced an oscillation frequency of 16 Hz throughout the measurement period of approximately 45 minutes. The sampling frequency of the accelerometer was approximately 195 Hz, so that a maximum frequency of about 95 Hz can be detected. Usually, the first few seconds of the data set are discarded as transient oscillation. The data set without this initial stabilization phase was modeled in [15]. In the following, we analyze only that initial phase, which we defined to consist of the first 1500 accelerometer values (i.e., of the initial approximately 7.7 seconds). Apart from the main frequency, multiples of 8 Hz with small amplitudes can be expected to occur as a consequence of the sampling of the originally continuous-time phenomenon and due to the physical properties of the shaker table. We modeled this signal content by means of the truncated Fourier series

$$y_t = \frac{a_0}{2} + \sum_{j=1}^{12} a_j \cos(2\pi f_j x_t) + b_j \sin(2\pi f_j x_t) + e_t, \qquad (t = 1, \ldots, n) \tag{23}$$

with fixed frequencies $f_j = j \cdot 8$ Hz; the unknown Fourier coefficients a_0, a_1, ..., a_{12} and b_1, ..., b_{12}, are collected within the parameter vector $\boldsymbol{\xi}$. Concerning the colored noise, we specified on the one hand a time-variable AR(p)-process using the global polynomials x^0, x^1, ..., x^q, and on the other hand a time-constant AR(p)-process (which constitutes the special case $q = 0$ of the preceding model). We tried different autoregressive and polynomial model orders, beginning with $p = q = 1$, and identified the least orders for which the estimated, decorrelation-filtered residuals $\hat{\mathbf{u}}$ (obtained as the values $\mathbf{u}^{(i+1)}$ after convergence of the ECME algorithm) pass a periodogram-based white noise test. We used for this purpose the MATLAB routine `periodogram` to compute the onesided periodogram I_1, ..., I_M, which values give the normalized cumulated periodogram

$$S_0 = 0, \quad S_i = \frac{\sum_{k=1}^{i} I_k}{\sum_{k=1}^{M} I_k} \quad (i = 1, \ldots, M),$$

where M is the lower integer of $n/2$. The test compares the maximum cumulated periodogram excess $T = \max_i |S_i - i/M|$ over a cumulated, theoretical white noise periodogram with $1 - \alpha$ significance bounds (cf. Sect. 7.3.3 in [33]). We thus obtained as the most parsimonious colored noise description a time-variable AR(6) model with cubic polynomials ($q = 3$) and a time-constant AR(21) model (see Fig. 1 for the depiction of the two periodogram tests and Fig. 2 for the estimated AR coefficients). The adjustment involving the time-variable AR

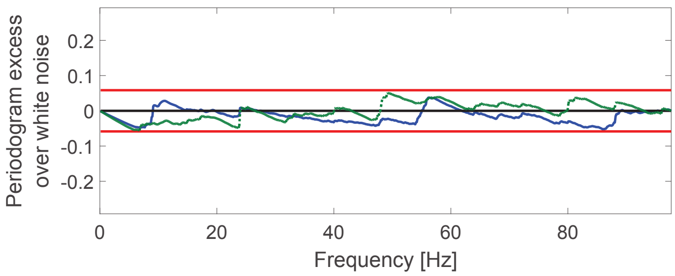

Fig. 1. Excess of the estimated periodogram of the decorrelated residuals for the time-variable AR(6) model (blue) and for the stationary AR(21) model (green) with respect to the theoretical white noise periodogram (black) and 99% significance bounds (red).

model yields for the estimated degree of freedom $\hat{\nu} = 4.8$ (indicating a rather heavy-tailed t-distribution), whereas we obtained the Gaussian limit $\hat{\nu} \to \infty$ for the time-constant model. The difference between these two models in terms of adjusted observations $\mathbf{A}\hat{\boldsymbol{\xi}}$ is also clearly discernible (see Fig. 3). Whereas the time-variable model reproduces the eventual oscillation amplitude quite accurately, much of the oscillation signal is absorbed into the colored noise residuals of the time-constant model. We therefore conclude that it is more reasonable to interpret and model the initial measurement phase, where the oscillation amplitude changes greatly before reaching a stable value, as a combination of outliers (leading to heavy tails) and non-stationary autocorrelation patterns interacting

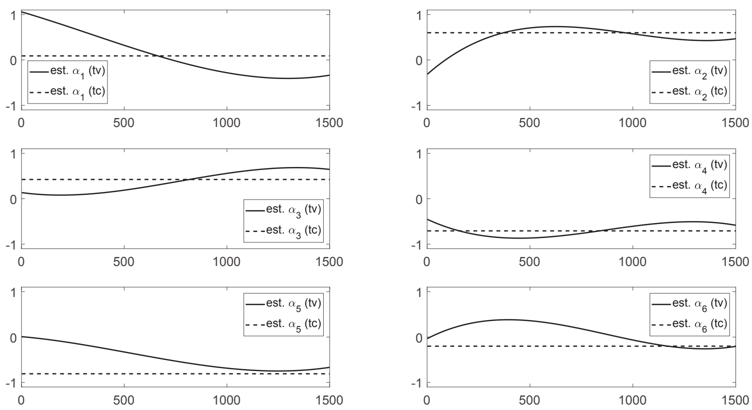

Fig. 2. Estimated coefficients of the time-variable (tv) AR(6) model and of the first six coefficients for the time-constant (tc) AR(21) model over the 1500 time instances.

with the Fourier series model. These patterns can be displayed as a time-variable power spectral density (see Fig. 4), defined by (cf. [35])

$$PSD(f,t) = \frac{\hat{\sigma}_u^2}{\left|1 - \sum_{j=1}^{p} \hat{\alpha}_{j,t} e^{-i2\pi jf}\right|^2}, \tag{24}$$

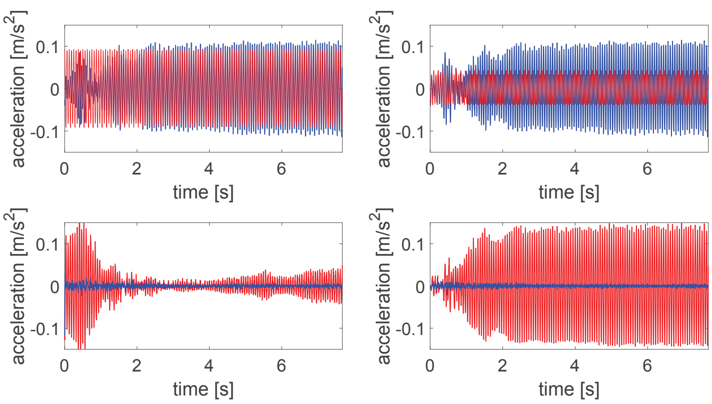

Fig. 3. Top row: plot of the complete dataset $\mathbf{y}$ (blue), of the adjusted observations involving the time-variable (tv) AR(6) model (in red on the left subplot), and of the adjusted observations $\mathbf{A}\hat{\boldsymbol{\xi}}$ involving the time-constant (tc) AR(21) model (in red on the right subplot). Bottom row: plots of the corresponding estimated residuals (decorrelation-filtered residuals $\hat{\mathbf{u}}$ in blue, colored noise residuals $\hat{\mathbf{e}}$ in red).

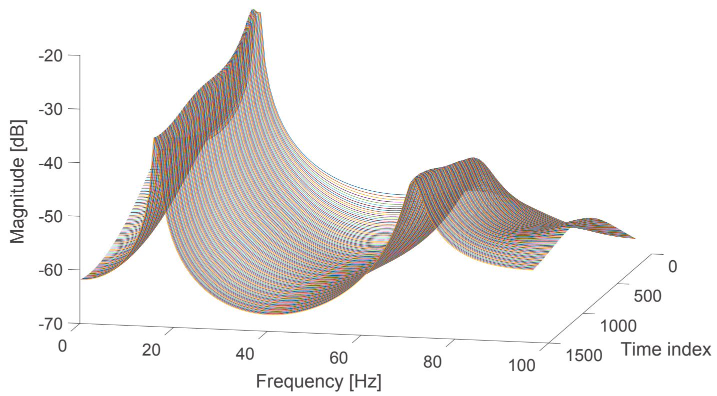

Fig. 4. Power spectral density based on the time-variable AR(6) processes.

where the standard deviation of the t-distributed white noise components is related to the estimated scale factor and degree of freedom via $\hat{\sigma}_u^2 = \frac{\hat{\nu}}{\hat{\nu}-2}\hat{\sigma}^2$. The evident fact that the PSDs have peaks around 16 Hz demonstrates that the oscillation signal is still partially captured by the colored noise model. As both the Fourier and the AR model have relationships with the frequency domain, this kind of interaction appears to be unavoidable.

Acknowledgments. The presented application of the PCB Piezotronics accelerometer within the vibration analysis experiment was performed as a part of the collaborative project "Spatio-temporal monitoring of bridge structures using low cost sensors" with ALLSAT GmbH, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) and the Central Innovation Programme for SMEs (ZIM Kooperationsprojekt, ZF4081803DB6). In addition, the authors would like to acknowledge the Institute of Concrete Construction (Leibniz Universität Hannover) for providing the shaker table and the reference accelerometer used within this experiment.

References

1. Azrak, R., Mélard, G.: AR Models with Time-Dependent Coefficients - a Comparison Between Several Approaches. Technical Report 0642, IAP Statistics Network, Interuniversity Attraction Pole (2006)
2. Bilmes, J.A.: A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov models. International Computer Science Institute, Berkeley/California, http://lasa.epfl.ch/teaching/lectures/ML_Phd/Notes/GP-GMM.pdf (1989)
3. Christmas, J., Everson, R.: Robust Autoregression: Student-t Innovations Using Variational Bayes. IEEE Transactions on Signal Processing 59, 48-57 (2011)

4. Dahlhaus, R.: Fitting Time Series Models to Nonstationary Processes. The Annals of Statistics 25, 1-37 (1997)
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society (Series B) 39, 1-38 (1977)
6. Eom, K.B.: Analysis of Acoustic Signatures from Moving Vehicles Using Time-Varying Autoregressive Models. Multidimensional Systems and Signal Processing 10, 357-378 (1999)
7. Francq, C., Gautier, A.: Large Sample Properties of Parameter Least Squares Estimates for Time-Varying ARMA Models. Journal of Time Series Analysis 25, 765-783 (2004)
8. Grenier, Y.: Time-Dependent ARMA Modeling of Nonstationary Signals. IEEE Transactions in Acoustics, Speech, and Signal Processing ASSP-31, 899-911 (1983)
9. Gupta, M.R., Chen, Y.: Theory and Use of the EM Algorithm. Foundations and Trends in Signal Processing 4, 223-296 (2011)
10. Härmä, A., Juntunen, M., Kaipio, J.P.: Time-Varying Autoregressive Modeling of Audio and Speech Signals. In: Proceedings of the 10th European Signal Processing Conference, Tampere, Finland, 2037-2040 (2000)
11. Hall, M.G., Oppenheim, A.V., Willsky, A.S.: Time-Varying Parametric Modeling of Speech. Signal Processing 5, 267-285 (1983)
12. Hallin, M.: Mixed Autoregressive-Moving Average Multivariate Processes with Time-Dependent Coefficients. Journal of Multivariate Analysis 8, 567-572 (1978)
13. Hamilton, J.D.: Time series analysis. Princeton University Press (1994)
14. Hildreth, C.: Asymptotic Distribution of Maximum Likelihood Estimators in a Linear Model with Autoregressive Disturbances. The Annals of Mathematical Statistics 40, 583-594 (1969)
15. 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, https://doi.org/10.1007/s00190-017-1062-6 (2017)
16. Kitagawa, G., Gersch, W.: A Smoothness Priors Time-Varying AR Coefficient Modeling of Nonstationary Covariance Time Series. IEEE Transactions on Automatic Control 30, 48-56 (1985)
17. Koch, K.R., Kargoll, B.: Expectation Maximization Algorithm for the Variance-Inflation Model by Applying the t-Distribution. Journal of Applied Geodesy 7, 217-225 (2013)
18. Lange, K.L., Little, R.J.A., Taylor, J.M.G.: Robust Statistical Modeling Using the t-Distribution. Journal of the American Statistical Association 84, 881-896 (1989)
19. Liporace, L.A.: Linear Estimation of Nonstationary Signals. The Journal of the Acoustical Society of America 58, 1288-1295 (1975)
20. Lira, I., Wöger, W.: Comparison Between the Conventional and Bayesian Approaches to Evaluate Measurement Data. Metrologia 43, S249-S259 (2006)
21. Liu, C.H., Rubin, D.B.: ML Estimation of the t Distribution using EM and its Extensions, ECM and ECME. Statistica Sinica 5, 19-39 (1995)
22. Liu, C.H., Rubin, D.B.: The ECME Algorithm: a Simple Extension of EM and ECM with Faster Monotone Convergence. Biometrika 81:633-648 (1994)
23. Luo, X., Mayer, M., Heck, B.: Analysing Time Series of GNSS Residuals by Means of AR(I)MA Processes. In: VII Hotine-Marussi Symposium on Mathematical Geodesy, International Association of Geodesy Symposia 137, pp. 129-134, Springer, Berlin, Heidelberg (2012)

24. McDonald, J.B.: Partially Adaptive Estimation of ARMA Time Series Models. International Journal of Forecasting 5, 217-230 (1989)
25. McDonald, J.B., Newey, W.K.: Partially Adaptive Estimation of Regression Models via the Generalized t Distribution. Econometric Theory 4, 428-457 (1988)
26. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. John Wiley & Sons, Hoboken, New Jersey (2008)
27. Meng, X., Rubin, D.B.: Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. Biometrika 80, 267-278 (1993).
28. Nassar, S., El-Sheimy, N.: Accuracy Improvement of Stochastic Modeling of Inertial Sensor Errors. Zeitschrift für Geodäsie, Geoinformation und Landmanagement 130, 146-155 (2005)
29. Park, M., Gao, Y.: Error and Performance Analysis of MEMS-based Inertial Sensors with a Low-cost GPS Receiver. Sensors 8, 2240-2261 (2008)
30. Rudoy, D., Quatieri, T.F., Wolfe, P.J.: Time-Varying Autoregressions in Speech: Detection Theory and Applications. IEEE Transactions in Acoustics, Speech, and Signal Processing 19, 977-989 (2011)
31. Schuh, W.D.: The Processing of Band-limited Measurements; Filtering Techniques in the Least Squares Context and in the Presence of Data Gaps. Space Science Reviews 108, 67-78 (2003)
32. Sanubari, J., Tokuda, K.: Non-Stationary Spectral Estimation Based on Robust Time Varying AR Model Excited by a t-Distribution Process. In: Proceedings of the IEEE Conference on Speech and Image Technologies for Computing and Telecommunications (TENCON'97), 51-54 (1997)
33. Schlittgen, R., Streitberg, B.H.J.: Zeitreihenanalyse. 9th edition. R. Oldenbourg Verlag, Munich (2001)
34. Subba Rao, T.: The Fitting of Non-Stationary Time-Series Models with Time-Dependent Parameters. Journal of the Royal Statistical Society (Series B) 32, 312-322 (1970)
35. Tary, J.B., Herrera, R.H., van der Baan, M.: Time-Varying Autoregressive Model for Spectral Analysis of Microseismic Experiments and Long-Period Volcanic Events. Geophys. J. Int. 196, 600-611 (2014)
36. Tsatsanis, M.K., Giannakis, G.B.: Time-Varying System Identification and Model Validation Using Wavelets. IEEE Transactions on Signal Processing 41, 3512-3523 (1993)
37. Wang, K., Xiong, S., Li, Y.: Modeling with Noises for Inertial Sensors. In: Position Location and Navigation Symposium (PLANS) 2012, IEEE/ION, 625-632 (2012)