Predicting in multivariate incomplete time series. Application of the expectation-maximisation algorithm supplemented by the Newton-Raphson method

Adam Korczyński^a

Abstract. Statistical practice requires various imperfections resulting from the nature of data to be addressed. Data containing different types of measurement errors and irregularities, such as missing observations, have to be modelled. The study presented in the paper concerns the application of the expectation-maximisation (EM) algorithm to calculate maximum likelihood estimates, using an autoregressive model as an example. The model allows describing a process observed only through measurements with certain level of precision and through more than one data series. The studied series are affected by a measurement error and interrupted in some time periods, which causes the information for parameters estimation and later for prediction to be less precise. The presented technique aims to compensate for missing data in time series. The missing data appear in the form of breaks in the source of the signal. The adjustment has been performed by the EM algorithm to a hybrid version, supplemented by the Newton-Raphson method. This technique allows the estimation of more complex models. The formulation of the substantive model of an autoregressive process affected by noise is outlined, as well as the adjustment introduced to overcome the issue of missing data. The extended version of the algorithm has been verified using sampled data from a model serving as an example for the examined process. The verification demonstrated that the joint EM and Newton-Raphson algorithms converged with a relatively small number of iterations and resulted in the restoration of the information lost due to missing data, providing more accurate predictions than the original algorithm. The study also features an example of the application of the supplemented algorithm to some empirical data (in the calculation of a forecasted demand for newspapers).

Keywords: missing data, multivariate time series, expectation-maximisation algorithm, Newton-Raphson algorithm

JEL: C13, C19, C61

1. Introduction

Data quality insight is one of the aspects of data science which supports data analysis by providing a framework that allows working with real data. Real data naturally tends to be erroneous, displaying measurement errors, incompleteness of data and other irregularities such as outliers.

The source of a measurement error can be different depending on the application. When analysing environmental and technical problems, it can result from instrument imprecision and differences in the locations of the assessments as shown in a pollution study by Butland et al. (2013). An example of a measurement error in economic analysis is provided by Fukuda (2005) in a study of a business cycle indicator affected by an error resulting from irregular sampling, which arose from

^a Warsaw School of Economics, Institute of Statistics and Demography, ul. Madalińskiego 6/8, 02-513 Warsaw, Poland, e-mail: akorczy@sgh.waw.pl, ORCID: https://orcid.org/0000-0002-1533-7197.

the need to produce business reports within a short period in order to evaluate the current business conditions. Ghassemi et al. (2015) discuss the issue of noisy data in medical studies where errors result from difficulties related to data collection. In effect, only irregularly sampled heterogeneous clinical records are available.

The problem of missing data has been extensively studied and described in statistical literature. A comprehensive overview of the topics related to the analysis of missing data is provided by Little & Rubin (2002). One of the recent works on missing data and other aspects of data quality and their implications from a survey perspective is by Laaksonen (2018).

The reason behind data incompleteness is twofold. It can be related to data collection issues, such as data entry delays or to the attitude of the respondents who may not be willing to provide certain information. While the former is usually less frequent and driven by completely random events, the latter can result in bias. Depending on the missing data mechanism, the missingness can lead to bias if the probability of observing an outcome would be determined by some subject characteristics or the outcome level itself. Naturally, a reduced amount of information causes the estimation of parameters to be imprecise. Various weighting adjustments and imputation techniques, including multiple imputation, have been implemented to address the above-mentioned problems in different data structures.

One other aspect in data quality assessment is the existence of outliers. Outlying observations can appear in univariate distribution, but when analysed from a multivariate perspective, they are likely to become a more complex problem. The analysis primarily focuses on influential observations, involving such values which affect the estimates or their standard errors. In some cases the outliers are simply erroneous data points and therefore one would expect a correction of such values to be made prior to the analysis. In practice, many erroneous data points can be detected directly by predefined edit checks. If an outlier turns to be a true value or it is not possible to clean the data for operational reasons, e.g. reporting time requirements, robust methods and models are designed to solve such problems. In univariate settings, the natural choice would be to utilise quantiles of the distribution and this concept can be extended to quantile regression models showing robustness against the observations that behave unlike most of the other ones (Koenker, 2005).

The aim of this paper is to assess the extension of the numeric technique for time series analysis outlined in Shumway & Stoffer (1982) in order to address the prediction problem, considering such aspects of data quality as measurement errors and data incompleteness. The model applied for smoothing and forecasting is the Kalman filter (Kalman, 1960), which is estimated through the expectation-maximisation (EM) algorithm. The standard EM algorithm requires closed-form solutions which are not readily available for more complex data structures. This in practice leads to the introduction of constraints imposed upon selected parameters

in order to simplify the numeric problem. The extension suggested in this paper involves incorporating the Newton-Raphson method within the regular EM algorithm in order to allow the estimation of parameters which are otherwise set to zero. This solution offers flexibility and reduces constraints when making assumptions for the analysed process, e.g. we can assume that there is a non-zero correlation between the measurement errors in a bivariate time series. The text provides a description of the studied process which uses an autoregressive model with random noise affecting the observed time series data.

2. Model for noisy time series

Kalman (1960) introduced a model for the description of processes with the aim of detecting signals in the presence of random noise. The detection process involves separating the signal from random noises, thereby providing basis for predicting the signals. The model finds its application when the observation is performed independently by two or more individuals or devices, with the goal to measure the same characteristic, although the data from each source are subject to measurement error. An example of such observation is described by Cajner et al. (2019). The purpose of the study was to combine two individual sources of data on the labour market (drawn from separate surveys) in order to reduce a measurement error which was in fact affecting both sources, and thus improve the accuracy of the estimates related to labour market characteristics, including the number of active and paid employees.

In multivariate time series settings, all data sources are utilised to predict the examined signal (e.g. the level of a specific characteristic), but in order to make that prediction, the data must first be cleaned from noise. In the notation, x_t will be the underlying, true level of the analysed parameter over time t = 1, 2, ..., n. Only y_t – the noisy data are observed. A process involving multiple data sources measuring the same characteristic with only limited precision can be expressed as follows (Shumway & Stoffer, 1982, p. 254):

$$\mathbf{y}_t = \mathbf{M}_t \, \mathbf{x}_t + \mathbf{v}_t \quad \text{for} \quad t = 1, 2, \dots, n, \tag{1}$$

where

 $\mathbf{y}_t - k \times 1$ is the vector of the observed series at time *t*;

- $\mathbf{M}_t k \times p$ is the design matrix transforming the unobserved $p \times 1$ vector \mathbf{x}_t into \mathbf{y}_t . If applied in incomplete data, \mathbf{M}_t would refer to the missing data indicator as defined in Little & Rubin (2002, Section 1.2), pointing to those elements of observation vector \mathbf{y}_t which are known and which are not;
- $\mathbf{v}_t k \times 1$ is the vector of the process error terms following normal distribution $N(\mathbf{\mu}, \mathbf{R})$ with zero-mean vector and $k \times k$ covariance matrix \mathbf{R} .

Let us assume that we want to obtain the measurement of \mathbf{x}_t , but in fact this variable is measured by two independent sources and we are not able to observe it directly. What we have is 2×1 vectors of measurements from each of the sources over time. Based on (1) the process is then as follows:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} x_t + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \text{ for } t = 1, 2, ..., n,$$
 (2)

where

$$\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} r_1^2 & r_{12} \\ r_{12} & r_2^2 \end{bmatrix} \right).$$
 (3)

Vector \mathbf{v}_t can be considered as a component aiming to capture the measurement error.

For example, we can assume that there are two sources of data for the same series. In economics it could be demand assessed by two individual sources. In medical statistics the application of the above concept would relate to models for laboratory measurements based on different types of samples (e.g. serum and plasma). Assuming the existence of a correlation between the two specimens (see for example Carey et al., 2016), the two series can be combined in order to describe the changes in the laboratory parameters under study over time.

We further specify the process describing random series x_t as a first-order autoregressive process, thereby introducing a correlation between the adjacent observations, i.e. the outcome at time t would depend on its level at time t - 1. As the model would simplify the true process, we include the term w_t to reflect deviations of the actual outcome from the modelled series. The model specification is then as follows:

$$\mathbf{x}_t = \mathbf{\Phi}_t \, \mathbf{x}_{t-1} + \mathbf{w}_t \quad \text{for} \quad t = 1, 2, \dots, n, \tag{4}$$

where

 $\Phi_t - p \times p$ is the transition matrix expressing the relationship between the adjacent values of the series over time;

 $\mathbf{w}_t - p \times 1$ is the vector of model error terms following a normal distribution, $N(\mathbf{\mu}, \mathbf{Q})$ is the zero-mean vector and $p \times p$ is covariance matrix \mathbf{Q} for the uncorrelated process. The initial value of process x_0 is assumed to be a normal random vector from $N(\mathbf{\mu}_0, \mathbf{\Sigma}_0)$, where $\mathbf{\Sigma}_0$ is a $p \times p$ covariance matrix. In the simplest case with univariate series, the model would take the following form:

$$x_t = \phi x_{t-1} + w_t$$
 for $t = 1, 2, ..., n$, (5)

where ϕ represents the autocorrelation between the adjacent elements of the series and where the variance of **w**_t is univariate and noted by q^2 .

Error \mathbf{w}_t captures the uncertainty displayed by the model, which reflects the ability to describe the process creating a series of observations.

In order to obtain the estimates of $x_1, x_2, ..., x_n$, observed series of vectors $y_1, y_2, ..., y_n$ are smoothed by using the model specified in (1) and (4). The same framework can be applied to calculate forecast $X_{n+1}, X_{n+2}, ..., X_{n+l}$, with *l* representing the number of periods for which the prediction is produced.

The process applicable to bivariate series which begins with an observation, and then proceeds to the smoothing of the series, on to the prediction is depicted in Figure 1.

t	Y_{1t}	Y_{2t}	-	X _t
1	y_{1t}	<i>y</i> ₂₁		<i>x</i> ₁
2	<i>y</i> ₁₂	y ₂₂		<i>x</i> ₂
÷	:	:		:
п	y_{1n}	y_{2n}	\rightarrow	x_n
				x_{n+1}
				x_{n+2} :
				x_{n+l}

Figure 1. Smoothing and predicting based on bivariate observed series

Source: author's work.

In the formulation of the model above, there are two random terms capturing the specific parts of the overall variability occurring in the process. Random term \mathbf{w}_t in (4) captures the uncertainty related to the model specification in relation to the actual process, while \mathbf{v}_t in (1) is to express the measurement error or the noise introduced by the measurement technique or displayed by the nature of the observed data.

The model parameters from (1) and (4) are estimated based on an iterative version of the maximum likelihood approach. The usage of the EM algorithm supplemented by Newton-Raphson allows expanding the model complexity, also taking the practical perspective into account, including handling missing data. The following sections describe the application of the estimating algorithm to obtain the maximum likelihood estimates of the model for bivariate incomplete data.

3. Estimation of the model

3.1. The likelihood function and the estimating algorithm

Assuming that the sample is large and the observation units are selected independently, the model specified by (1) and (4) can be estimated using the maximum likelihood method. The method maximises the joint density expressed as a product of individual probabilities given the population distribution (for details on the method see for example Mittelhammer, 2013, Section 8.3; Little & Rubin, 2002, Chapter 6). For the model specified in Section 1 the likelihood is a product of three factors related to the initial condition for the process, the autoregressive part and the part explaining how underlying series x_t (t = 1, 2, ..., n) are transformed into actual observations (Shumway & Stoffer, 2017, p. 306):

$$f(\mathbf{y}|\boldsymbol{\theta}) = f(\mathbf{x}_0|\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \prod_{t=1}^n f(\mathbf{x}_t|\mathbf{x}_{t-1}, \boldsymbol{\Phi}, \mathbf{Q}) \prod_{t=1}^n f(\mathbf{y}_t|\mathbf{x}_t, \mathbf{R}),$$
(6)

where θ represents the set of parameters describing the population of interest and requiring estimation.

Given joint density (6), the log-likelihood assuming $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$, considering $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n$, and ignoring the constants is as follows (Shumway & Stoffer, 1982, p. 256):

$$\ln L \left(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y}\right) = -\frac{1}{2} \ln |\boldsymbol{\Sigma}_{0}| - \frac{1}{2} (\mathbf{x}_{0} - \boldsymbol{\mu}_{0})' \boldsymbol{\Sigma}_{0}^{-1} (\mathbf{x}_{0} - \boldsymbol{\mu}_{0}) - \frac{n}{2} \ln |\mathbf{Q}| - \frac{1}{2} \sum_{t=1}^{n} (\mathbf{x}_{t} - \boldsymbol{\Phi} \mathbf{x}_{t})' \mathbf{Q}^{-1} (\mathbf{x}_{t} - \boldsymbol{\Phi} \mathbf{x}_{t}) - \frac{n}{2} \ln |\mathbf{R}| - \frac{1}{2} \sum_{t=1}^{n} (\mathbf{y}_{t} - \mathbf{M}_{t} \mathbf{x}_{t})' \mathbf{R}^{-1} (\mathbf{y}_{t} - \mathbf{M}_{t} \mathbf{x}_{t}).$$
(7)

The maximum likelihood (ML) estimates of the model given by (1) and (4) are obtained by maximising the log-likelihood function (7) with respect to parameters μ_0 , Σ_0 , Φ , Q and R. In relation to simple problems, the ML estimators can be derived directly from the log-likelihood function. When the level of complexity increases, which may result from the nature of the series (e.g. higher dimensions, more complex model specifications), the direct maximisation is rarely available and in practice referring to optimisation techniques such as Newton-Raphson or EM algorithms becomes inevitable.

Shumway and Stoffer (1982, pp. 256–257) provide formulas for estimating the parameters of (1) and (4) using the EM algorithm. This approach enables addressing two types of data irregularities:

- the expectation step allows the implementation of the smoothing estimator so that the observed data (e.g. two-dimensional series) assessing the same effect but with some level of a measurement error represented by **R** can be transformed into the smoothed series;
- the expectation step can be further extended to incorporate compensation for the missing data, i.e. finding the maximum likelihood estimates through an iterative process replacing the missing values by their expectations drawn from a conditional distribution from a prespecified population.

The algorithms outlined in Shumway & Stoffer (1982) as well as in Little & Rubin (2002, Chapter 8) focus on applications assuming a direct approach to the maximisation step. The suggestion proposed in this paper is to combine the EM algorithm with the Newton-Raphson method, in this case operating as a sub-algorithm. Some examples of such an approach have been described as hybrid maximisation methods in Little & Rubin (2002, pp. 186–188) (see for example Lange, 1995).

The goal of the approach proposed in this paper is to allow more complex model specifications, which is basically equivalent to relaxing some of the constraints used to simplify the computations. In Shumway and Stoffer (1982), it is the assumption on uncorrelated measurement errors in vector \mathbf{v}_t which provides a simplified version of the **R** matrix with a covariance occurring between the measurement errors equal to zero.

The EM algorithm formulation is depicted in Korczyński, 2018, Chapter 4; Little & Rubin, 2002, Chapter 8, and Molenberghs & Kenward, 2007, pp. 93–103. The algorithm consists of an initial step and two main steps: the expectation step, in which the expected value of the log-likelihood is calculated, and the maximisation step, in which updates to the model parameters are found so that they maximise the expected likelihood at the current iteration. In the presented version of the algorithm, we assume that the actual series of y_t (t = 1, 2, ..., n) is not fully observable, and then data vector **y** is split into observed y^{obs} and missing y^{mis} . The iterative process consists of three steps.

Initial step. Set the parameter vector to certain initial values $\mathbf{\Theta}^{(0)}$. In a missing data application the complete-case or available-case approach can be used (for details see for example Little & Rubin, 2002, Chapter 8) to obtain the initial estimates.

Expectation step. Calculate the expected value of the log-likelihood function with the current estimates of parameter vector $\mathbf{\theta}^{(i)}$:

$$Q(\mathbf{\theta}|\mathbf{\theta}^{(i)}) = E[\ln L(\mathbf{\theta}|\mathbf{y})|\mathbf{y}^{obs},\mathbf{\theta}^{(i)}].$$
(8)

In practice, this step requires finding the expected values of the sufficient statistics and replacing the unknown components by their expected values. The unknown components may be related to the smoothing of the series to identify the underlying unobserved process (Shumway & Stoffer, 1982, Section 2) or to the missing values (Dempster et al., 1977, pp. 3–4).

Maximisation step. Find $\mathbf{\theta}^{(i+1)}$ maximising the log-likelihood function considering the current expectations over the log-likelihood function:

$$Q(\mathbf{\theta}^{(i+1)}|\mathbf{\theta}^{(i)}) \ge Q(\mathbf{\theta}^{(i+1)}|\mathbf{\theta}^{(i)}) \text{ for all } \mathbf{\theta}.$$
(9)

After having updated $\mathbf{\Theta}^{(i+1)}$, we need to go back to the first step and the cycle is repeated until convergence is reached, i.e. until the changes observed for the estimates in question are lower than an arbitrarily small value ε .

The EM algorithm, as specified above, ensures convergence to the maximum of the log-likelihood (Molenberghs & Kenward, 2007, p. 95), although the rate of the convergence can be slow.

The suggested extension as outlined in Korczyński (2018, pp. 216–225), involves the replacement of the maximisation step with the Newton-Raphson maximisation of the parameters for which closed-form solutions are not available. In the model outlined by (1) and (4), the Newton-Raphson step would entail estimating the elements of variance-covariance matrix **R**, depicting the behaviour of the measurement error terms. The general notation of the maximisation is as follows:

$$\widehat{\mathbf{\theta}}^{(i+1)} = \widehat{\mathbf{\theta}}^{(i)} - \left\{ E\left[\frac{\partial^2 \ln L(\mathbf{\theta}^{(i)} | \mathbf{y})}{\partial \mathbf{\theta}^{(i)} \partial \mathbf{\theta}^{(i)}} \right] \right\}^{-1} E\left[\frac{\partial \ln L(\mathbf{\theta}^{(i)} | \mathbf{y})}{\partial \mathbf{\theta}^{(i)}} \right].$$
(10)

In a regular framework which ignores the noise around the signal and any potentially missing data we would proceed directly to the maximisation of the log-likelihood (7). However, we assume that the true process is only approximated by individual data series and we let some elements of the series be missing, which is unavoidable in statistical practice. An example of this process which will be further discussed in this text is outlined in Figure 2.

t	Y_{1t}	Y_{2t}	_	X_t
1	y_{11}	<i>y</i> ₂₁		<i>x</i> ₁
2	y_{12}	y_{22}		x_2
3		y_{23}		x_3
4		y_{24}		x_4
5	y_{15}	<i>y</i> ₂₅		x_5
6	y_{16}		\rightarrow	x_1
	:	:		:
n	y_{1n}	•	_	x_n
				x_{n+1}
				x_{n+2}
				:
				x_{n+l}

Figure 2. Smoothing and predicting based on bivariate observed series with missing data

Firstly, we need to find the underlying series through the smoothing of the observed data $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ to receive $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$.

Secondly, we need to consider the fact that some parts of the observed series are missing. Both aspects are dealt with by determining the expected value of the log-likelihood function (7).

After calculating the maximum likelihood estimates, we can proceed to the prediction of the next l elements of the series.

In order to simplify the notation, we assume that the actual process is observed through two data sources as described by (2) and (5). In this case the log-likelihood can be written as:

$$\ln L \left(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y} \right) = -\frac{1}{2} \ln \sigma_0^2 - \frac{1}{2\sigma_0^2} (x_0 - \mu_0)^2 - \frac{n}{2} \ln q^2 - \frac{1}{2q^2} \sum_{t=1}^n (x_t - \phi x_{t-1})^2 - \frac{n}{2} \ln \left| \begin{matrix} r_1^2 & r_{12} \\ r_{12} & r_2^2 \end{matrix} \right| - \frac{1}{2} \sum_{t=1}^n \left(\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} x_t \right)' \begin{bmatrix} r_1^2 & r_{12} \\ r_{12} & r_2^2 \end{bmatrix}^{-1} \left(\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} - \begin{bmatrix} m_{1t} \\ m_{2t} \end{bmatrix} x_t \right).$$
(11)

The log-likelihood function (11) can be expressed as:

$$\ln L \left(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y}\right) = -\frac{1}{2} \ln \sigma_0^2 - \frac{1}{2\sigma_0^2} (x_0^2 - 2\mu_0 x_0 + \mu_0^2) - \frac{n}{2} \ln q^2 - \frac{1}{2q^2} \sum_{t=1}^n (x_t^2 - 2\phi x_t x_{t-1} + \phi^2 x_{t-1}^2) - \frac{n}{2} \ln(r_1^2 r_2^2 - r_{12}^2) - \frac{1}{2(r_1^2 r_2^2 - r_{12}^2)} \sum_{t=1}^n [(y_{1t}^2 - 2m_{1t} y_{1t} x_t + m_{1t}^2 x_t^2) r_2^2 - 2(y_{1t} y_{2t} - m_{2t} y_{1t} x_t - m_{1t} y_{2t} x_t - m_{1t} m_{2t} x_t^2) r_{12} + (y_{2t}^2 - 2m_{2t} y_{2t} x_t + m_{2t}^2 x_t^2) r_1^2].$$
(12)

The implementation of the Newton-Raphson step allows assuming the existence of a correlation between the error terms in the y_{1t} and y_{2t} observed series.

In order to apply the EM algorithm, we firstly specify the expectation of loglikelihood (8). This step requires calculating:

- the expectation of the unobserved x₁, x₂, ..., x_n series. This allows for the smoothing
 of the observed process to assess the underlying series;
- the expectation of the incomplete elements of the observed **y**₁, **y**₂, ..., **y**_n series. This part addresses the issue of missing data in the process.

The expected log-likelihood for (12) can be expressed as:

$$E[\ln L (\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})] = -\frac{1}{2} \ln \sigma_0^2 - \frac{1}{2\sigma_0^2} [E(x_0^2) - 2\mu_0 E(x_0) + \mu_0^2] - \frac{n}{2} \ln q^2 - \frac{1}{2q^2} \sum_{t=1}^n [E(x_t^2) - 2\phi E(x_t x_{t-1}) + \phi^2 E(x_{t-1}^2)] - \frac{n}{2} \ln (r_1^2 r_2^2 - r_{12}^2) - \frac{1}{2(r_1^2 r_2^2 - r_{12}^2)} \sum_{t=1}^n \{ [E(y_{1t}^2) - 2m_{1t} E(y_{1t} x_t) + m_{1t}^2 E(x_t^2)] r_2^2 - 2[E(y_{1t} y_{2t}) - m_{2t} E(y_{1t} x_t) - m_{1t} E(y_{2t} x_t) - m_{1t} m_{2t} E(x_t^2)] r_{12} + [E(y_{2t}^2) - 2m_{2t} E(y_{2t} x_t) + m_{2t}^2 E(x_t^2)] r_1^2 \}.$$
(13)

The next step is to find the expectations of \mathbf{x}_t and \mathbf{y}_t in (13). The estimator of the \mathbf{x}_t smoothed series is depicted in Subsection 3.2. The calculation of the expectations of the missing elements of \mathbf{y}_t is presented in Subsection 3.3. Given the expected \mathbf{x}_t and \mathbf{y}_t at a specific iteration of the algorithm, the log-likelihood function is maximised with respect to the $\mathbf{\theta}^i$ parameters of interest to update the estimates and obtain $\mathbf{\theta}^{i+1}$, which is then utilised to recalculate the expectations of the log-likelihood. The maximisation is outlined in Subsection 3.4. The algorithm runs in cycles, from the expectation step to the maximisation step and over again, until convergence occurs.

3.2. Smoother estimator

For the specified autoregressive model accounting for noise given by (1) and (4), we utilise the Kalman smoother estimator taking relevant steps needed to estimate the studied parameters. The process is outlined in Shumway & Stoffer (1982, pp. 262–263). The expected value of \mathbf{x}_t is noted as:

$$\mathbf{x}_t^n = E[\mathbf{x}_t | \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n], \tag{14}$$

while the error variance-covariance matrix takes the following form:

$$\mathbf{P}_t^n = E[(\mathbf{x}_t - \mathbf{x}_t^n)(\mathbf{x}_t - \mathbf{x}_t^n)'|\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n].$$
(15)

As noted in Anderson & Moore (1979, pp. 37–38), error variance-covariance matrix (15) measures how effective the $\hat{\mathbf{x}}_t^n$ estimate is. The trace of \mathbf{P}_t^n from (15) is given by:

A. KORCZYŃSKI Predicting in multivariate incomplete time series. Application of the expectation... 27

$$tr \mathbf{P}_{t}^{n} = E\{tr[(\mathbf{x}_{t} - \mathbf{x}_{t}^{n})(\mathbf{x}_{t} - \mathbf{x}_{t}^{n})'|\mathbf{y}_{1}, \mathbf{y}_{2}, \dots, \mathbf{y}_{n}]\}$$
$$E\{\|\mathbf{x}_{t} - \mathbf{x}_{t}^{n}\|^{2}|\mathbf{y}_{1}, \mathbf{y}_{2}, \dots, \mathbf{y}_{n}\},$$
(16)

and it is the conditional error variance for $\hat{\mathbf{x}}_t^n$. The optimisation involves finding $\hat{\mathbf{x}}_t^n$, for which the error variance is minimised.

For t = 1, 2, ..., n and with the initial condition for the process of $\mathbf{x}_0^0 = \boldsymbol{\mu}_0$ and $\mathbf{P}_0^0 = \boldsymbol{\Sigma}_0$, the smoothing is performed by means of the following set of recursive equations:

$$\mathbf{x}_{t}^{t-1} = \mathbf{\Phi} \mathbf{x}_{t-1}^{t-1},\tag{17}$$

$$\mathbf{P}_t^{t-1} = \mathbf{\Phi} \mathbf{P}_{t-1}^{t-1} \mathbf{\Phi}' + \mathbf{Q},\tag{18}$$

$$\mathbf{K}_t = \mathbf{P}_t^{t-1} \mathbf{M}_t' (\mathbf{M}_t \mathbf{P}_t^{t-1} \mathbf{M}_t' + \mathbf{R})^{-1}, \tag{19}$$

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t^{t-1}),$$
(20)

$$\mathbf{P}_t^t = \mathbf{P}_t^{t-1} + \mathbf{K}_t \mathbf{M}_t \mathbf{P}_t^{t-1}.$$
 (21)

To find estimates $\hat{\mathbf{x}}_t^n$ in (14) and $\hat{\mathbf{P}}_t^n$ in (15), we carry out backward calculations for t = n, n - 1, ..., 1 on:

$$\mathbf{J}_{t-1} = \mathbf{P}_{t-1}^{t-1} \mathbf{\Phi}' (\mathbf{P}_t^{t-1})^{-1},$$
(22)

$$\hat{\mathbf{x}}_{t-1}^{n} = \mathbf{x}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{x}_{t}^{n} - \mathbf{\Phi}\mathbf{x}_{t-1}^{t-1}),$$
(23)

$$\widehat{\mathbf{P}}_{t-1}^{n} = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}(\mathbf{P}_{t}^{n} - \mathbf{P}_{t}^{t-1})\mathbf{J}_{t-1}^{\prime}.$$
(24)

In expected log-likelihood (13), we need to calculate the expected value of the product of the subsequent elements of unobserved series $E(x_t x_{t-1})$. This requires the covariance of the elements, which is calculated for t = n, n - 1, ..., 2 using

$$\widehat{\mathbf{P}}_{t-1,t-2}^{n} = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-2}' + \mathbf{J}_{t-1} \big(\mathbf{P}_{t,t-1}^{n} - \mathbf{\Phi} \mathbf{P}_{t-1}^{t-1} \big) \mathbf{J}_{t-2}',$$
(25)

where

$$\mathbf{P}_{n,n-1}^{n} = (\mathbf{I} - \mathbf{K}_{n} \mathbf{M}_{n}) \mathbf{\Phi} \mathbf{P}_{t-1}^{t-1}.$$
(26)

The expected values of \mathbf{x}_t are calculated using formulas (17)–(26) recursively. The formulas required to calculate the expected values of the missing elements of \mathbf{y}_t are presented in the next subsection.

3.3. Expectations of incomplete data

3.3.1. General notation for multivariate normal distribution

In order to find the expectations of \mathbf{y}_t for the incomplete series, a reference must be made to the properties of the multivariate normal distribution. The concept is described in Korczyński (2018, pp. 135–137). Under the parametric assumption stating that the process is normal, we can describe it through mean vector $\mathbf{\mu}$ and variance-covariance matrix $\mathbf{\Sigma}$. The process starts from separating the complete and incomplete part of vector \mathbf{y}_t so that the complete part is followed by the missing elements. In effect, the arrangement of elements in the mean vector and variancecovariance matrix takes the form shown in (27), where $\mathbf{\mu}_1$ and $\mathbf{\Sigma}_1$ represent the known elements for a specific missing data pattern (solid line), while $\mathbf{\mu}_2$ and $\mathbf{\Sigma}_2$ refer to the missing data elements (dashed line):

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_2 \end{bmatrix}.$$
(27)

For example, let us consider the three-element vector \mathbf{y}_t at time t = 4 in Figure 3. The observation vector is given by $\mathbf{y}_t = [\mathcal{Y}_{14} \quad \mathcal{Y}_{34}]'$ with the middle element missing. The necessary rearrangement would result in $\mathbf{y}_t^* = [\mathcal{Y}_{14} \quad \mathcal{Y}_{34} \quad \cdot]'$, which is equivalent to $\mathbf{y}_t^{(1)} = [\mathcal{Y}_{14} \quad \mathcal{Y}_{34}]'$ and $\mathbf{y}_t^{(2)}$ missing,¹ with the respective changes to mean vector and variance-covariance matrix (27): $\mathbf{\mu}_1 = [\mu_1 \quad \mu_3]'$, $\mathbf{\mu}_2 = \mu_2$, $\mathbf{\Sigma}_1 = \begin{bmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{13} & \sigma_3^2 \end{bmatrix}'$, $\mathbf{\Sigma}_2 = \sigma_2^2$, $\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{21}' = [\sigma_{12} \quad \sigma_{23}]'$. The joint mean vector and variance-covariance matrix for that missing data pattern can be expressed as:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_3 & \mu_2 \end{bmatrix}', \\ \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{13} & \sigma_{12} \\ \sigma_{13} & \sigma_3^2 & \sigma_{23} \\ \sigma_{12} & \sigma_{23} & \sigma_2^2 \end{bmatrix}.$$
(28)

¹ Notation $\mathbf{y}_t^{(1)}$ is to represent the observed part of vector \mathbf{y}_t at time t. Similarly, $\mathbf{y}_t^{(2)}$ is to represent the missing part. Indices 1 and 2 can refer to various combinations of the three elements of \mathbf{y}_t , and therefore it should be noted that y_{1t} (the value of the first element of \mathbf{y}_t at time t) and $\mathbf{y}_t^{(1)}$ represent different notations.

	Da	ata			Re	sulting missi	ng data patt	tern
t	Y_{1t}	Y_{2t}	Y_{2t}		S	<i>Y</i> _{1s}	Y_{2s}	Y_{3s}
1	<i>y</i> ₁₁	<i>y</i> ₂₁		\rightarrow	1	0	0	Х
2	y_{12}	y_{22}			2	0	х	х
3	y_{13}	•			3	0	х	0
4	<i>y</i> ₁₄		y ₃₄	-				

Figure 3. Example of a missing data pattern

Source: Korczyński (2018, p. 133).

In the *i*-th iteration of the EM algorithm the expected value of missing part $\mathbf{y}_t^{(2)}$ is calculated as follows:

$$E\left(\mathbf{y}_{t}^{(2)} \middle| \mathbf{y}_{t}^{(1)}, \widehat{\mathbf{\mu}}^{(i-1)}, \widehat{\mathbf{\Sigma}}^{(i-1)}\right) = \widehat{\mathbf{\alpha}}_{2|1}^{(i)} + \widehat{\mathbf{\beta}}_{2|1}^{(i)} \mathbf{y}_{t}^{(1)},$$
(29)

where:

$$\widehat{\boldsymbol{\alpha}}_{2|1}^{(i)} = \widehat{\boldsymbol{\mu}}_{2}^{(i-1)} - \widehat{\boldsymbol{\Sigma}}_{21}^{(i-1)} \widehat{\boldsymbol{\Sigma}}_{11}^{-1(i-1)} \widehat{\boldsymbol{\mu}}_{2}^{(i-1)},$$
(30)

$$\widehat{\beta}_{2|1}^{(i)} = \widehat{\Sigma}_{21}^{(i-1)} \widehat{\Sigma}_{11}^{-1(i-1)}.$$
(31)

The variance-covariance matrix of the error term in the model specified by (29) is given by

$$\widehat{\mathbf{\Sigma}}_{2|1}^{(i)} = \widehat{\mathbf{\Sigma}}_{22}^{(i-1)} - \widehat{\mathbf{\Sigma}}_{21}^{(i-1)} \widehat{\mathbf{\Sigma}}_{11}^{-1(i-1)} \widehat{\mathbf{\Sigma}}_{12}^{(i-1)}.$$
(32)

In order to calculate conditional expected values $E(y_{jt}^2|w)$ and $E(y_{jt}y_{kt}|w)$ for $j \neq k$, and with *w* representing the condition, a reference is made to the properties of the variance-covariance matrix (see Härdle & Simar, 2015, pp. 123–125):

$$Cov(\mathbf{x}, \mathbf{y}) = E(\mathbf{x}\mathbf{y}') - E(\mathbf{x})[E(\mathbf{y})]'.$$
(33)

The expectations are conditional on the current parameter estimates and the observed part of vector y_t .

We can further adjust (33) to the desired notation using $\mathbf{y}_t^{(1)}$ and $\mathbf{y}_t^{(2)}$:

$$E\left[\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)'\right] = E\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\left[E\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\right]' + Cov\left[\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\right],$$
(34)

where $Cov\left[\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\left(\mathbf{y}_{t}^{(2)}|\mathbf{y}_{t}^{(1)}\right)\right] = \boldsymbol{\Sigma}_{2|1}$ is the variance-covariance matrix of the error terms in model (29).

With regard to the earlier example presented in Figure 3, we now consider vector \mathbf{y}_t at time t = 3. In this case missing data vector $\mathbf{y}_t^{(2)}$ has two elements representing y_{23} and y_{33} . The expected values of $E(y_{2t}^2|w)$, $E(y_{3t}^2|w)$ and $E(y_{2t}y_{3t}|w)$ can be calculated using formula (34). The respective equation can be expressed as:

$$\begin{bmatrix} E(y_{2t}^{2}|w) & E(y_{2t}y_{3t}|w) \\ E(y_{2t}y_{3t}|w) & E(y_{3t}^{2}|w) \end{bmatrix} =$$

$$= \begin{bmatrix} E(y_{2t}|w)^{2} & E(y_{2t}|w)E(y_{3t}|w) \\ E(y_{2t}|w)E(y_{3t}|w) & E(y_{3t}|w)^{2} \end{bmatrix} + \Sigma_{2|1},$$
(35)

where

$$\boldsymbol{\Sigma}_{2|1} = \begin{bmatrix} \sigma_{y_{2t}|y_{1t}}^2 & \sigma_{y_{2t}y_{3t}|y_{1t}} \\ \sigma_{y_{2t}y_{3t}|y_{1t}} & \sigma_{y_{3t}|y_{1t}}^2 \end{bmatrix}.$$
 (36)

From the property presented in (34) we can derive well-known equations defining variance and covariance through expected values:

$$D^2 X = E X^2 - (E X)^2, (37)$$

and

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$
(38)

3.3.2. Adjusting for missing values in an autoregressive model accounting for noise

The calculation of the expected values of \mathbf{y}_t required for the model outlined by (2) and (5) involve the derivation of the mean vector and variance covariance matrix (27) specified for that modelling framework.

For the case of bivariate \mathbf{y}_t and univariate \mathbf{x}_t , which is equivalent to the notation given by (2) and (5), the mean vector of \mathbf{y}_t is given by

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E(y_{1t}) \\ E(y_{2t}) \end{bmatrix} = \begin{bmatrix} E(m_{1t}x_t + v_{1t}) \\ E(m_{2t}x_t + v_{2t}) \end{bmatrix} = \begin{bmatrix} E(m_{1t}x_t) \\ E(m_{2t}x_t) \end{bmatrix} = \begin{bmatrix} m_{1t}\hat{x}_t^n \\ m_{2t}\hat{x}_t^n \end{bmatrix}, \quad (39)$$

for t = 1, 2, ..., n and based on the fact that the expected value of the error terms in **v**_t equals zero (3), and where \hat{x}_t^n (23) is the estimate of x_t^n (14), calculated through

the described recursive set of equations. In order to provide the probabilistic part of the distribution specification, i.e. the elements of Σ , we should first note that the variance of \mathbf{y}_t is

$$\begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \end{bmatrix} = \begin{bmatrix} D^2(y_{1t}) \\ D^2(y_{2t}) \end{bmatrix} = \begin{bmatrix} D^2(m_{1t}x_t + v_{1t}) \\ D^2(m_{2t}x_t + v_{2t}) \end{bmatrix}$$

$$= \begin{bmatrix} m_{1t}^2 D^2(x_t) + r_1^2 \\ m_{2t}^2 D^2(x_t) + r_2^2 \end{bmatrix} = \begin{bmatrix} m_{1t}^2 \hat{P}_t^n + r_1^2 \\ m_{2t}^2 \hat{P}_t^n + r_2^2 \end{bmatrix},$$

$$(40)$$

for t = 1, 2, ..., n, where r_1^2 and r_2^2 are the variances of error term \mathbf{v}_t in (3). The following equation illustrates the use of the property of variance of the sum of random variables (see for example Jóźwiak & Podgórski, 2006, p. 107):

$$D^{2}(X+Y) = D^{2}(X) + D^{2}(Y) + 2Cov(X,Y),$$
(41)

and assuming that two random variables $m_{1t}x_t$ and v_{1t} are independent, which is equivalent to $Cov(m_{1t}x_t, v_{1t}) = 0$. Variance \hat{P}_t^n is calculated using (24).

According to model specification (2), y_{1t} and y_{2t} are strongly linearly correlated. In fact, the two series describe the same process and the differences result from measurement errors. By that, even with a large discrepancy between variances $D^2(y_{1t})$ and $D^2(y_{2t})$, the Pearson correlation coefficient between y_{1t} and y_{2t} is approximately equal to one ($\rho \approx 1$). This leads to the following approximation of the covariance between y_{1t} and y_{2t} :

$$\sigma_{12} = Cov(y_{1t}, y_{2t}) \approx D(y_{1t})D(y_{2t}) = \sigma_1 \sigma_2,$$
(42)

which comes from a well-known formula for calculating the Pearson correlation coefficient (see for example Jóźwiak & Podgórski, 2006, p. 107):

$$\rho = \frac{Cov(X,Y)}{D(X)D(Y)}.$$
(43)

Formulas (40) and (42) provide the following specification of variance-covariance matrix Σ of the joint distribution of \mathbf{y}_t elements in (27):²

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}. \tag{44}$$

 $^{^2}$ At this stage μ and Σ are still prior to the rearrangement for the observed and missing parts.

Based on (39) and (44) we can proceed to the description of the steps needed to calculate the expected values of the missing elements of \mathbf{y}_t conditional on the observed data and the current estimates of the mean vector and the variance-covariance matrix at a given iteration of the EM algorithm, denoted as $\boldsymbol{\mu}^{(1i)}$ and $\boldsymbol{\Sigma}^{(i)}$, where *i* is the iteration number.

We further assume that the observed time series is as shown in Figure 2, i.e. for some parts of the series one of the components can be missing for a number of periods. For the parts of the series where y_{1t} is missing, the mean vector and variance-covariance matrix (27), after undergoing a rearrangement into the observed and incomplete part, take the form of:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_2 \\ \mu_1 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_2^2 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{bmatrix}.$$
(45)

and the expected value of y_{1t} is calculated using (29)

$$E(y_{1t}|w) = E(y_{1t}|y_{2t}, \hat{\boldsymbol{\mu}}^{(i-1)}, \hat{\boldsymbol{\Sigma}}^{(i-1)}) = \hat{\boldsymbol{\alpha}}_{2|1}^{(i)} + \hat{\boldsymbol{\beta}}_{2|1}^{(i)} \boldsymbol{y}_{t}^{(1)}$$

$$= \hat{\mu}_{1}^{(i-1)} + \frac{\sigma_{12}^{(i-1)}}{\sigma_{2}^{2}^{(i-1)}} (y_{2t} - \hat{\mu}_{2}^{(i-1)}), \qquad (46)$$

where *w* represents the conditional term in $E(y_{1t}|y_{2t}, \hat{\mu}^{(i-1)}, \hat{\Sigma}^{(i-1)})$ and is applied here to simplify the further notation.

Expected value $E(y_{1t}^2|w)$ is calculated using (36) and in the considered case the expression is simplified to

$$E(y_{1t}^2|w) = E(y_{1t}^2|y_{2t}, \widehat{\boldsymbol{\mu}}^{(i-1)}, \widehat{\boldsymbol{\Sigma}}^{(i-1)}) = E(y_{1t}|w)^2.$$
(47)

Similarly, for the missing y_{2t} the mean vector and variance-covariance matrix (27) take the following form:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix},$$
(48)

and this leads to the expected value of y_{2t} :

A. KORCZYŃSKI Predicting in multivariate incomplete time series. Application of the expectation... 33

$$E(y_{2t}|w) = E(y_{2t}|y_{1t}, \hat{\mu}^{(i-1)}, \hat{\Sigma}^{(i-1)}) = \hat{\alpha}_{2|1}^{(i)} + \hat{\beta}_{2|1}^{(i)} \mathbf{y}_{t}^{(2)}$$

$$= \hat{\mu}_{2}^{(i-1)} + \frac{\sigma_{12}^{(i-1)}}{\sigma_{1}^{2}^{(i-1)}} (y_{1t} - \hat{\mu}_{1}^{(i-1)}).$$
(49)

For expected value $E(y_{2t}^2|w)$ we obtain

$$E(y_{2t}^2|w) = E(y_{2t}^2|y_{1t}, \hat{\boldsymbol{\mu}}^{(i-1)}, \hat{\boldsymbol{\Sigma}}^{(i-1)}) = E(y_{2t}|w)^2.$$
(50)

The expectations of \mathbf{x}_t determined to smooth the time series, and \mathbf{y}_t to adjust for the missing data described in the preceding subsections are used to calculate the expected log-likelihood function (13), which can be rewritten to

$E[\ln L\left(\boldsymbol{\theta}|\mathbf{x},\mathbf{y}\right)] =$

where l_1 would depend on the missing data pattern. For a complete \mathbf{y}_t , we note l_1^{obs} :

$$l_{1}^{obs} = \sum_{t=1}^{n} [y_{1t}^{2} - 2m_{1t}y_{1t}E(x_{t}) + m_{1t}^{2}E(x_{t}^{2})]r_{2}^{2} -2[y_{1t}y_{2t} - m_{2t}y_{1t}E(x_{t}) - m_{1t}y_{2t}E(x_{t}) - m_{1t}m_{2t}E(x_{t}^{2})]r_{12} + [y_{2t}^{2} - 2m_{2t}y_{2t}E(x_{t}) + m_{2t}^{2}E(x_{t}^{2})]r_{1}^{2}.$$
(52)

For an incomplete \mathbf{y}_t , its missing elements in l_1 are replaced by their expected values according to (46)-(47) and (49)-(50).

3.4. Maximisation of the log-likelihood

The EM algorithm involves the maximisation of the log-likelihood function (51) with respect to the parameters of the model given by (2) and (5): μ_0 , σ_0^2 , ϕ , q^2 and **R**. The closed-form solutions maximising the log-likelihood can be found for most of the parameters. As we do not assume the error terms in (2) to be uncorrelated, covariance r_{12} leads to the respective log-likelihood being specified as a higher-order polynomial. For the estimation of that parameter, the Newton-Raphson step is proposed as an iterative alternative to direct maximisation.

In order to find both the closed-form and the Newton-Raphson solutions, we first calculate the partial derivatives of the log-likelihood with respect to the model parameters

Przegląd Statystyczny. Statistical Review 2021 | 1

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})}{\partial \theta_j}\right],\tag{53}$$

where $\boldsymbol{\theta} = [\mu_0 \quad \sigma_0^2 \quad \boldsymbol{\phi} \quad q^2 \quad r_1^2 \quad r_2^2 \quad r_{12}]'$. For $\mu_0, \sigma_0^2, \boldsymbol{\phi}, q^2$, we can directly maximise the log-likelihood by equating the first derivative of the log-likelihood (53) to zero:

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})}{\partial \theta_j}\right] = 0.$$
 (54)

Therefore, from ① in (51) we follow with the equations below:

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})}{\partial \mu_0}\right] = 0,$$
(55)

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial \sigma_0^2}\right] = 0,$$
(56)

through which we obtain the following maximum likelihood estimates of μ_0 , σ_0^2 :

$$\hat{\mu}_0 = \hat{x}_0^n,\tag{57}$$

$$\hat{\sigma}_0^2 = \hat{P}_0^n,\tag{58}$$

where $\hat{\chi}_0^n$ and \hat{P}_0^n are from (23) and (24).

The two parameters defining the autoregressive process are estimated according to the formula resulting from 0 in (51):

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial \phi}\right] = 0,$$
(59)

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial q^2}\right] = 0.$$
 (60)

With the use of properties described in (37) and (38), the estimators take the form of:

$$\hat{\phi} = \frac{\sum_{t=1}^{n} \hat{x}_{t}^{n} \hat{x}_{t-1}^{n} + \hat{P}_{t,t-1}^{n}}{\sum_{t=1}^{n} (\hat{x}_{t-1}^{n})^{2} + \hat{P}_{t-1}^{n}},\tag{61}$$

$$\hat{q}^2 = \sum_{t=1}^n (\hat{x}_t^n - \phi \hat{x}_{t-1}^n)^2 + \hat{P}_t^n - \phi \left(2\hat{P}_{t,t-1}^n + \phi \hat{P}_{t-1}^n\right),\tag{62}$$

where \hat{x}_t^n and \hat{x}_{t-1}^n are from (23), \hat{P}_t^n and \hat{P}_{t-1}^n are given by (24), and $\hat{P}_{t,t-1}^n$ is provided by (25). The variances of error terms r_1^2 and r_2^2 from ③ in (51) can be maximised directly from:

34

A. KORCZYŃSKI Predicting in multivariate incomplete time series. Application of the expectation... 35

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial r_1^2}\right] = 0,$$
(63)

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})}{\partial r_2^2}\right] = 0.$$
(64)

First, let us observe that ③ in (51) can be rewritten as:

$$-\frac{n}{2}\ln(r_1^2r_2^2 - r_{12}^2) - \frac{1}{2(r_1^2r_2^2 - r_{12}^2)}\sum_{t=1}^n ar_2^2 - 2br_{12} + cr_1^2,$$
(65)

where: $\begin{aligned} a &= y_{1t}^2 - 2m_{1t}y_{1t}E(x_t) + m_{1t}^2E(x_t^2), \\ b &= y_{1t}y_{2t} - m_{2t}y_{1t}E(x_t) - m_{1t}y_{2t}E(x_t) - m_{1t}m_{2t}E(x_t^2), \\ c &= y_{2t}^2 - 2m_{2t}y_{2t}E(x_t) + m_{2t}^2E(x_t^2). \end{aligned}$

After taking the first derivative with respect to the parameter of interest, in this case r_1^2 , and multiplying the whole expression by denominator $(r_1^2 r_2^2 - r_{12}^2)^2$, we note that r_1^2 would be cancelled out by the second element of the sum and the resulting maximum likelihood estimator would take the form of

$$\hat{r}_1^2 = -\frac{1}{nr_2^2} \left[nr_{12}^2 + \left(\sum_{t=1}^n ar_2^2 - 2br_{12} + c\frac{r_{12}^2}{r_2^2} \right) \right].$$
(66)

Similarly, we obtain the following for r_2^2 :

$$\hat{r}_2^2 = -\frac{1}{nr_1^2} \left[nr_{12}^2 + \left(\sum_{t=1}^n cr_1^2 - 2br_{12} + a\frac{r_{12}^2}{r_1^2} \right) \right].$$
(67)

For the estimation of r_{12} , we introduce the Newton-Raphson step, which operates within the EM algorithm according to (10):

$$\hat{\mathbf{r}}_{12}^{i+1} = \hat{\mathbf{r}}_{12}^{i} - \left\{ E\left[\frac{\partial^2 \ln L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})}{\partial r_{12}^{(i)} \partial r_{12}^{(i)}} \right] \right\}^{-1} E\left[\frac{\partial \ln L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y})}{\partial r_{12}^{(i)}} \right],$$
(68)

where

$$E\left[\frac{\partial \ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial r_{12}^{(l)}}\right] = \frac{nr_{12}}{r_1^2 r_2^2 - r_{12}^2} - \frac{r_{12}}{(r_1^2 r_2^2 - r_{12}^2)^2} \sum_{t=1}^n ar_2^2 + cr_1^2 + \frac{1}{r_1^2 r_2^2 - r_{12}^2} \sum_{t=1}^n b, \quad (69)$$

and

$$E\left[\frac{\partial^{2}\ln L(\boldsymbol{\theta}|\mathbf{x},\mathbf{y})}{\partial r_{12}^{(i)}\partial r_{12}^{(i)}}\right] = \frac{n(r_{12}^{2}+r_{1}^{2}r_{2}^{2})}{(r_{1}^{2}r_{2}^{2}-r_{12}^{2})^{2}} - \frac{3r_{12}^{2}+r_{1}^{2}r_{2}^{2}}{(r_{1}^{2}r_{2}^{2}-r_{12}^{2})^{3}}\sum_{t=1}^{n}ar_{2}^{2} + cr_{1}^{2} + \frac{2r_{12}(r_{12}^{2}+3r_{1}^{2}r_{2}^{2})}{(r_{1}^{2}r_{2}^{2}-r_{12}^{2})^{3}}\sum_{t=1}^{n}b.$$
(70)

The Newton-Raphson step in the EM procedure actually operates as a subalgorithm, finding the maximum likelihood estimate of r_{12} iteratively, at each run of the EM. It is worth noting that we can apply the use of the Newton-Raphson algorithm to more than one parameter within the EM, which extends the usage of this estimation procedure.

To summarise, the EM algorithm with the Newton-Raphson step operates according to the following scheme (compare with Shumway & Stoffer, 1982, p. 258):

- set the initial values of parameters μ_0 , σ_0^2 , ϕ , q^2 , r_1^2 , r_2^2 and r_{12} ;
- for the current values of the estimates, calculate the expected values of x_t using xⁿ_t and xⁿ_{t-1} from (23), Pⁿ_t and Pⁿ_{t-1} from (24), and Pⁿ_{t,t-1} from (25) and y_t using (46)-(37) and (49)-(50) for the two missing data patterns, respectively;
- estimate the parameters of the substantive model using formulas $\hat{\mu}_0$, $\hat{\sigma}_0^2$, $\hat{\phi}$, \hat{q}^2 , \hat{r}_1^2 , \hat{r}_2^2 and \hat{r}_{12} (57), (58), (61), (62), (66), (67) and (68), noting that (68) for r_{12} is the iterative approach. If the estimates of the parameters are not stabilised, move to step 2, whereas if the convergence has been reached, stop the process. The estimates from the last iteration of the EM algorithm are the maximum likelihood estimates of the model described by (2) and (5).

Assessment and application of the EM algorithm with the Newton-Raphson step

4.1. Assessment of the algorithm

The EM algorithm with the Newton-Raphson step (EMNR) as described in the previous section has been applied to the data generated from a distribution imitating the process of autoregression affected by noise, with two observed series measuring the same, unobserved underlying process. The algorithm has been applied using a programme code written in SAS[®] IML, available upon request.

The sample of n = 200 values was drawn from a population described by two equations – (2) and (5). The parameters of the model from which the sample was drawn are shown in Table 1. The values of the parameters have been selected so as to demonstrate an exemplary autoregressive process with a relatively strong autocorrelation between adjacent elements of time series. The goal was to compare the estimates with the actual values.

μ_0	σ_0^2	ϕ	q^2	r_{1}^{2}	r_{2}^{2}	<i>r</i> ₁₂
47,000	8,319,328.444	0.98	4,000	4,000	4,000	2,000
Source: autho	or's work					

Table 1. Parameters of the model used to generate the sample

Source: author's work

At first, the initial value of process x_0 was drawn from $N(\mu_0, \sigma_0)$. Subsequently, the autoregressive parameter was set equal to $\phi = 0.98$. Further, an assumption was made that the observed series y_t for t = 1, 2, ..., n can differ by the error terms, and the transition factor from the unobserved to observed series is one, which is equivalent to $\mathbf{M}_t = \begin{bmatrix} 1 & 1 \end{bmatrix}'$. Having specified ϕ , the values of x_t were drawn from the autoregressive process with the w_t disturbance term. Following that and having specified \mathbf{M}_t , observed series \mathbf{y}_t were drawn from $\mathbf{R} = \begin{bmatrix} 4000 & 2000 \\ 2000 & 4000 \end{bmatrix}$. The draw of the series was followed by the removal of observations from y_{1t} for t between 51 and 84 and t > 191, and for y_{2t} for t between 140 and 167. This was to imitate the situation of losing one source of the signal for a specific time period. Overall, this resulted in r = 133 complete observations (i.e. 66.5% of data were complete).

Figure 4 shows the series of the generated x_t values (circles). The breaks in the series represent the time periods with a limited signal, which in this case involved observing only one element – y_t for some time.





Source: author's work.





Figure 5. Iterations history for the EMNR algorithm

Source: author's work.

A. KORCZYŃSKI Predicting in multivariate incomplete time series. Application of the expectation... 39

The EMNR was applied with the number of iterations equal to 10,000. In order to speed up the processing, parameter r_{12} was set to 0 for the initial 20 iterations, which allowed stabilising at first r_1^2 , r_2^2 . From the 21st iteration onwards, r_{12} was approximated at each iteration.

The high number of iterations allows observing the behaviour of the algorithm for a longer period. In fact, a far smaller number of iterations was required to reach convergence (see Figure 5). The estimates of x_t for t = 1, 2, ..., n after smoothing and adjusting for incomplete data are indicated with pluses in Figure 4. By comparing the pluses with the curve showing the actual process, we can see that it was resumed regardless of the missing data which occurred for some periods in both series.

Furthermore, the descriptive statistics for the Absolute Percentage Error (APE) were calculated following an estimation using the EMNR and the standard EM algorithm, in which $r_{12} = 0$. The Mean Absolute Percentage Error (MAPE) was calculated according to the following formula:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{x_t - \hat{x}_t}{x_t} \right|,$$
(71)

with the result values multiplied by a 100.

The results are presented in Table 2. We can see that the MAPE expressing the relative deviation from the actual values *ex post* is smaller for the EMNR than for the standard EM algorithm (0.84 vs 0.92) and display a lower standard deviation of APE (1.172 vs 1.283). The prediction error is on average less than 0.9% of the actual values of the underlying process. The forecasts from both algorithms were compared using the Diebold-Mariano (DM) test (<u>1995</u>) with a forecast horizon of three periods, through a one-sided test with the null hypothesis of no difference and an alternative that the EMNR forecast is more accurate than the EM forecast. The DM yielded the test statistic of DM = -3.06 with the *p*-value = 0.001 and, thus, the null hypothesis was rejected at $\alpha = 0.05$. The predictions from the standard EM algorithm. The result is determined by the possibility to relax the assumption for $r_{12} = 0$, which is not required in the EMNR algorithm. The estimates of the standard EM algorithm are presented in the last row of Table 3.

The iteration process of the EMNR is shown in Table 3 and Figure 5, which demonstrate that the algorithm required approximately 1,000 iterations to reach convergence for all the estimated parameters, i.e. μ_0 , σ_0^2 , ϕ , q^2 , r_1^2 , r_2^2 and r_{12} . What is more, after reaching convergence the estimates remained stable, which is reflected in the curves representing the estimates at subsequent iterations (Figure 5) and in their comparison with the estimates from the last iterations (Table 3). Autoregressive

parameter estimate ϕ is equal to the parameter value. Estimates μ_0 , q^2 , r_2^2 and r_{12} are close to the true underlying values. A smaller value than expected is obtained for r_1^2 . A similar situation is with initial variance σ_0^2 , for which it would be the highest deviation from the actual value, however this last parameter is of less importance for the process description. The missing data would be mostly affecting the estimate of the variance of the first error term in (2).

The Newton-Raphson sub-algorithm was set to work in a loop of 50 iterations. The actual convergence was immediate, occurring after only several repetitions of the cycle. An example of sub-iteration for the 10th iteration of the EM algorithm is presented in Figure 6.

Table 2. Descriptive statistics of the APE for the smoothing estimator, and results of the DMtest for forecast comparison

Method	n	Mean	Std. Dev.	Minimum	Maximum
EMNR algorithm	200	0.84	1.172	0.0004	6.09
EM algorithm	200	0.92	1.283	0.0012	6.48

Note. A one-sided DM test with an alternative hypothesis stating that EMNR forecasts are more accurate than EM forecasts and a forecast horizon of three periods yielded a test statistic of DM = -3.06 with a *p*-value = 0.001.

Source: author's calculations. DM test generated using the 'multDM' R package.

Iteration	μ_0	σ_0^2	φ	q^2	r_{1}^{2}	r_{2}^{2}	<i>r</i> ₁₂
1	21,321.20	7,072.78	0.9847	6,863,171.56	1,809,904.50	1,813,848.31	0
2	21,346.27	7,066.46	0.9887	4,633,299.45	958,295.53	961,020.78	0
3	21,384.15	7,056.85	0.9896	4,079,657.86	536,900.08	539,165.70	0
4	21,428.38	7,045.62	0.9898	3,929,662.89	319,062.70	320,625.27	0
5	21,475.19	7,033.73	0.9899	3,871,281.88	197,721.15	198,743.50	0
:	:	:	:	:	:	:	:
21	22,271.12	6,831.60	0.9901	3,586,632.42	2,051.60	2,069.82	862.78
22	22,321.24	6,818.88	0.9901	3,572,790.50	2,074.15	2,097.57	902.68
23	22,371.36	6,806.15	0.9901	3,559,551.61	2,106.44	2,133.39	932.70
24	22,421.48	6,793.42	0.9901	3,546,362.07	2,132.28	2,163.73	959.99
25	22,471.60	6,780.69	0.9901	3,533,192.22	2,155.03	2,191.44	984.09
:	:	:	:	:	:	:	:
9,991	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.49	1,657.64
9,992	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,993	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,994	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,995	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,996	49,630.05	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,997	49,630.06	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,998	49,630.06	0.74	0.9804	5,006.55	2,496.28	3,217.48	1,657.64
9,999	49,630.06	0.74	0.9804	5,006.55	2,496.29	3,217.48	1,657.64
10,000	49,630.06	0.74	0.9804	5,006.55	2,496.29	3,217.48	1,657.64
10,000*	49640.3	0.81	0.9804	6,481.13	1,688.3	2,225.3	0

Table 3. Iteration process for the EMNR algorithm

*The last row includes the estimates from the standard EM algorithm for comparison with the EMNR. Source: author's calculations.





Source: author's calculations.

4.2. Forecasting demand based on incomplete bivariate time series data

The model given by (2) and (5) has been utilised to analyse the demand for newspapers and to create a forecast for the three months following the last observations to illustrate the application of the EMNR algorithm. The data have been drawn from a website providing the service of monitoring the sales and circulation of the press titles in Poland (Teleskop, n.d.). The analysed dataset consists of two series of the monthly sales and distribution of the printed Polish daily newspaper 'Rzeczpospolita' between January 2016 and May 2018 (Table 4). The sample size equals n = 29. Note that for practical applications a larger sample would be expected as the underlying estimation method is the maximum likelihood.

The goal of the example is to demonstrate the application of the EMNR algorithm in the context of a business problem using empirical data. The problem selected for analysis is experienced by newspaper companies and its aim is to predict the correct number of printouts required by particular selling points.

In order to illustrate the functioning of the algorithm, the values in the brackets in Table 4 were considered missing, which, as in sampled data description, is to mimic the break in the reception of one of the signals used to assess the demand in this case.

The programme written in SAS^{*} IML and available upon request has been applied to estimate the parameters of the model describing the evolution of the demand for the analysed newspaper. The number of iterations was set to 10,000. Unlike in the case of sampled data, the convergence is slower. The algorithm reached a stable level for the studied parameters close to the end of the predefined iteration number, which might have resulted from the fact that the available sample was very small. The estimates for the early and final iterations of the algorithm are shown in Table 5. The convergence of the Newton-Raphson sub-algorithm was immediate.

Figure 7 presents observed series \mathbf{y}_t along with the estimated smoothed series reflecting underlying process x_t . We can see that the model fit leads to a nearly linear evolution around the shifts captured by the error terms deviations.

Equations (17)–(21) have been used to calculate the forecast for the demand for the 'Rzeczpospolita' newspaper for the three months following the last observation (see Shumway & Stoffer, 1982, p. 262).

Predictions with the standard errors are shown in Table 6. Their values indicate a downward trend observed for the series in the analysis.





Source: author's calculations.

Time	Sales	Distribution
1	47,389	49,117
2	44,026	44,115
3	44,263	44,398
4	43,812	43,910
5	43,044	43,203
6	43,168	43,379
7	42,961	43,059
8	41,601	(41,694)
9	41,698	(41,813)
10	40,965	(41,071)
11	40,675	(40,780)
12	40,647	40,739
13	40,927	41,000
14	37,753	37,842
15	37,390	37,417
16	38,047	38,139
17	37,612	37,695
18	37,487	37,514
19	37,606	37,652
20	37,064	37,092
21	37,025	37,093
22	36,540	36,648
23	36,782	36,897
24	36,859	36,918
25	(34,080)	34,117
26	(34,191)	34,225
27	(35,058)	35,084
28	(34,297)	34,364
29	(34,156)	34,247

Table 4. Sales and distribution	of the 'Rzeczpospolita'	' newspaper between January	2016
and May 2018			

Source: author's calculations based on data from Teleskop (n.d.).

Table 5. Iteration history for the EM algorithm – newspaper dataset

Iteration	μ_0	σ_0^2	φ	<i>q</i> ²	r_{1}^{2}	r_{2}^{2}	<i>r</i> ₁₂
1	21,321.2	7,072.78	0.9978	27,289,942.3	2,221,952.9	2,498,100.3	0
2	21,327.8	7,071.02	0.9996	26,186,180.8	1,401,354.4	1,553,840.5	0
3	21,334.8	7,069.17	0.9996	26,164,134.9	904,320.5	1,008,221.3	0
4	21,341.9	7,067.29	0.9995	26,203,873.2	602,394.4	673,174.3	0
5	21,349.1	7,065.41	0.9995	26,224,958.2	413,633.3	463,387.0	0
:	:	:	:	:	:	:	:
21	21,464.3	7,035.09	0.9995	25,979,283.8	72,433.6	89,516.6	26,224.1
22	21,471.5	7,033.19	0.9996	25,851,652.3	72,421.1	95,489.9	30,575.4
23	21,478.7	7,031.28	0.9996	25,725,415.0	72,710.1	102,398.1	33,939.5
24	21,486.0	7,029.37	0.9996	25,599,466.9	72,555.9	108,888.4	37,129.5
25	21,493.2	7,027.44	0.9997	25,471,929.7	72,184.8	115,224.2	40,116.4
÷	÷	:	:	:	:	:	:
9,991	46,052.6	111.00	0.9893	767,016.6	61,832.4	311,668.8	138,820.5
9,992	46,052.7	110.98	0.9893	767,014.7	61,832.4	311,668.8	138,820.5
9,993	46,052.7	110.97	0.9893	767,012.9	61,832.4	311,668.8	138,820.5
9,994	46,052.8	110.95	0.9893	767,011.0	61,832.4	311,668.8	138,820.5
9,995	46,052.9	110.93	0.9893	767,009.1	61,832.4	311,668.8	138,820.5
9,996	46,052.9	110.92	0.9893	767,007.2	61,832.4	311,668.8	138,820.5
9,997	46,053.0	110.90	0.9893	767,005.3	61,832.4	311,668.8	138,820.5
9,998	46,053.1	110.89	0.9893	767,003.4	61,832.4	311,668.8	138,820.5
9,999	46,053.1	110.87	0.9893	767,001.5	61,832.4	311,668.8	138,820.5
10,000	46,053.2	110.86	0.9893	766,999.6	61,832.4	311,668.8	138,820.5

Source: author's calculations.

Time	\hat{x}_{t+1}	$\sqrt{\hat{P}_{t+1}}$
$t + 1 \dots t + 2 \dots t + 3 \dots t + 3 \dots t$	33,881 33,520 33,162	875 1,232 1,501

Table 6. Prediction of the demand for the 'Rzeczpospolita' newspaper for June-August 2018

Source: author's calculations.

5. Conclusions

Statistical practice is faced with the issue of handling various imperfections resulting from data's nature. Different types of measurement errors need to be modelled and irregularities in the observed data such as missing observations must be taken into account. The Kalman filter is one of the tools that allows modelling noisy time series data. The article focused on exploring the application of the model and the underlying estimation process to situations when empirical data series contain measurement errors and are incomplete. The incompleteness involves situations when one of the sources of the signal is broken for some time, leaving less precise information to estimate the parameters and make predictions. The technique presented in the text was built on the concept described by Shumway & Stoffer (1982), extending the algorithm from the paper to a hybrid version, including the Newton-Raphson sub-algorithm.

The extended version of the algorithm has been verified using sampled data from a model imitating the studied process. The verification showed that the EMNR converged with a relatively small number of iterations and produced stable maximum likelihood estimates. The estimation accounted for incompleteness of the observed data vector, restoring the missing information so that the estimates of the substantive model parameters converged towards the parameters of the datagenerating model. The extended EMNR algorithm provided statistically significantly more accurate predictions as compared to the standard EM algorithm. An application to empirical data has also been included jointly with a calculation of the demand for newspapers predicted in subsequent periods closely after the last observation.

The suggested extension of the algorithm could supplement the solution for obtaining maximum likelihood estimates with a better assessment of uncertainty resulting from missing data through multiple imputations within the Bayesian paradigm. The basic scheme for complete data estimation of an autoregressive process can be found in Geweke (2005, Section 7.1). The approach would require a further assessment of the probabilistic characteristics of the estimates, which would lead to a proper assessment of the confidence intervals.

References

- Anderson, B. D. O., & Moore, J. B. (1979). *Optimal Filtering*. Hoboken: Prentice-Hall. http://users.cecs.anu.edu.au/~john/papers/BOOK/B02.PDF.
- Butland, B. K., Armstrong, B., Atkinson, R. W., Wilkinson, P., Heal, M. R., Doherty, R. M., & Vieno, M. (2013). Measurement error in time-series analysis: a simulation study comparing modelled and monitored data. *BMC Medical Research Methodology*, 13, 1–12. https://doi.org /10.1186/1471-2288-13-136.
- Cajner, T., Crane, L. D., Decker, R. A., Hamins-Puertolas, A., & Kurz, C. (2019). Improving the accuracy of economic measurement with multiple data sources: The case of payroll employment data (NBER Working Paper No. 26033). https://doi.org/10.3386/w26033.
- Carey, R. N., Jani, C., Johnson, C., Pearce, J., Hui-Ng, P., & Lacson, E. (2016). Chemistry Testing on Plasma Versus Serum Samples in Dialysis Patients: Clinical and Quality Improvement Implications. *Clinical Journal of the American Society of Nephrology*, 11(9), 1675–1679. https://doi.org/10.2215/CJN.09310915.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599.
- Fukuda, K. (2005). Forecasting economic time series with measurement error. Applied Economics Letters, 12(15), 923–927. https://doi.org/10.1080/13504850500119161.
- Geweke, J. (2005). Contemporary Bayesian Econometrics and Statistics. Hoboken: John Wiley & Sons.
- Ghassemi, M., Pimentel, M. A. F., Naumann, T., Brennan, T., Clifton, D. A., Szolovits, P., & Feng, M. (2015). A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data. Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto. http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper /view/9393.
- Härdle, W. K., & Simar, L. (2015). Applied Multivariate Statistical Analysis (4th edition). Berlin, Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-662-45171-7.
- Jóźwiak, J., & Podgórski, J. (2009). *Statystyka od podstaw* (6th edition). Warszawa: Polskie Wydawnictwo Ekonomiczne.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering, 82(1), 35–45. https://doi.org/10.1115/1.3662552.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. https://doi.org/10.1017 /CBO9780511754098.
- Korczyński, A. (2018). Screening wariancji jako narzędzie wykrywania zmowy cenowej. Istota i znaczenie imputacji danych. Warszawa: Oficyna Wydawnicza SGH.
- Laaksonen, S. (2018). Survey Methodology and Missing Data: Tools and Techniques for Practitioners. Cham: Springer. https://doi.org/10.1007/978-3-319-79011-4.
- Lange, K. (1995). A quasi-Newton acceleration of the EM algorithm. *Statistica Sinica*, 5(1), 1–18. http://www3.stat.sinica.edu.tw/statistica/oldpdf/A5n11.pdf.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (2nd edition). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119013563.

- Mittelhammer, R. (2013). *Mathematical Statistics for Economics and Business* (2nd edition). New York: Springer. https://doi.org/10.1007/978-1-4614-5022-1.
- Molenberghs, G., & Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons. https://doi.org/10.1002/9780470510445.
- Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3(4), 253–264. https://doi.org/10.1111 /j.1467-9892.1982.tb00349.x.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th edition). Cham: Springer. https://doi.org/10.1007/978-3-319-52452-8.
- Teleskop. (n.d.). *Polskie badania czytelnictwa*. Retrieved July 14, 2018, from https://www.teleskop.org.pl /zkdp/index.jsp?p=news.