Stud. Univ. Babeş-Bolyai Math. 66(2021), No. 1, 179–196
DOI: 10.24193/subbmath.2021.1.15

Kálmán's filtering technique in structural equation modeling

Marianna Bolla and Fatma Abdelkhalek

Dedicated to Professor Gheorghe Moroşanu on the occasion of his 70th anniversary.

Abstract. Structural equation modeling finds linear relations between exogenous and endogenous latent and observable random vectors. In this paper, the model equations are considered as a linear dynamical system to which the celebrated R. E. Kálmán's filtering technique is applicable. An artificial intelligence is developed, where the partial least squares algorithm of H. Wold and the block Cholesky decomposition of H. Kiiveri et al. are combined to estimate the parameter matrices from a training sample. Then the filtering technique introduced is capable to predict the latent variable case values along with the prediction error covariance matrices in the test sample. The recursion goes from case to case along the test sample, without having to re-estimate the parameter matrices. The algorithm is illustrated on real life sociological data.

Mathematics Subject Classification (2010): 62H05, 62P25, 68T99.

Keywords: Structural equation modeling, linear dynamical systems, Kálmán's filtering, artificial intelligence, application to social sciences.

1. Introduction

We consider structural equation model (SEM) for two latent random vectors that depend through a linear model on two observable random vectors, respectively (they usually include exogenous and endogenous variables). This kind of models was first investigated by T. Haavelmo [1], who obtained the Nobel Prize for it later. Unlike the traditional factor analysis, where latent variables were introduced and given a meaning based on the factor loadings, here the latent variables are organic parts of the model. The latent variables, e.g., alienation, ambition in [2] or mobility in our example, are given by the experts, and the observed (measurement) variables are indicators of them. In this way, so-called inner and outer relations are stated between the latent variables and between the observable and latent ones, respectively.

The estimation of the parameter matrices of this model was elaborated both in the Gaussian and distribution-free cases, former by K. G. Jöreskog (LISREL) [2], while the other by H. Wold (PLS) [8] in the 1970-1980s. These two approaches are sometimes called covariance-based and component-based SEM. However, we can consider the model equations as a linear dynamical system to which the celebrated R. E. Kálmán's filtering technique [3] is applicable. This technique was developed in the 1960s for time series to make predictions for the hidden state variables of a state space model, and was used in the lunar landing, for instance. We will show how to apply this technique in the more complicated dynamical system, containing two state and two observable equations, describing inner relations between the observable and latent variables, both for the exogenous and endogenous ones. Our contribution is that we connect these two approaches.

The parameter matrices are estimated from a training sample. We combine the first stage of the PLS algorithm of H. Wold to estimate the case values of the latent variables and the method of H. Kiiveri et al. [5] to decompose the inverse of the product moment matrix obtained with the latent case value estimates. At this point, we apply the block Cholesky decomposition. Then the filtering technique to be introduced is capable to make predictions for the endogenous variables based on the exogenous ones, through the latent variables. The driving force is that we propagate the error covariance matrices of the exogenous and endogenous latent variables in a recursion.

The test sample is a succession of observations coming one by one (like a time series or just subsequent observations), and the algorithm predicts their endogenous variables based on their own exogenous ones. Our contribution is that we combine existing methods for parameter estimation, and then apply filtering technique for prediction, so we develop an artificial intelligence. The computational gain is that the parameter matrices need not be estimated for every new-coming case in the test sample, but are estimated only once, in the training sample. The method is distribution-free (just second moments are used in the linear state equations) and applicable to small training, and not necessarily independent test samples.

The organization of the paper is as follows. In Section 2, the most important notions and facts about best linear predictions in Hilbert spaces are introduced. In Section 3, the prediction and propagation of the error covariance matrices are derived in two stages. The main results are summarized in Theorem 3.1 of Section 3.3. Then the proposed algorithm is illustrated on real life sociological data in Section 4. Eventually, the last Section 5 discusses the benefits of the proposed method together with some possible further perspectives.

2. Preliminaries

The following linear dynamical system that resembles the one to which R. E. Kálmán gave a recursive algorithm is considered:

$$\boldsymbol{B}\boldsymbol{\eta} = \boldsymbol{A}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

where $\boldsymbol{\eta}$ is m- and $\boldsymbol{\xi}$ is n-dimensional latent vector, $\boldsymbol{B}$ and $\boldsymbol{A}$ are $m \times m$ and $m \times n$ coefficient matrices, and $\boldsymbol{\zeta}$ is a random vector of residuals of uncorrelated components.

It is also uncorrelated with $\boldsymbol{\xi}$, and $\boldsymbol{B}$ is nonsingular. In the recursive models, $\boldsymbol{B}$ is upper triangular, with 1s along its main diagonal.

Here $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are not observed, but instead, the p-dimensional $\mathbf{Y}$ and the q-dimensional $\mathbf{X}$ are observed such that

$$\mathbf{X} = \boldsymbol{C}\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \quad \mathbf{Y} = \boldsymbol{G}\boldsymbol{\eta} + \boldsymbol{\delta},$$

where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of measurement errors in $\mathbf{X}$ and $\mathbf{Y}$, respectively. They are uncorrelated with each other and $\boldsymbol{\zeta}$. Typically, $n \leq q$ and $m \leq p$.

For the estimation of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{G}$, and the covariance matrices of the errors, there is the LISREL algorithm of K. G. Jöreskog [2] (assuming multivariate Gaussian distribution of the measurement variables and large sample sizes) and component-wise SEM algorithms (not postulating normality and being able to treat small sample sizes), e.g., [7, 8], at our disposal.

In the first stage of his PLS algorithm, H. Wold [8] gives an iteration to find the case values of the latent variables. He states that this fixed point iteration converges. We use only this first stage to calculate the product moment estimate of the covariance matrix of the latent variables. Then we decompose the inverse of this matrix as $\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$ with $\boldsymbol{L}$ and $\boldsymbol{D}$ having the form

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{B}^T & \boldsymbol{O} \\ -\boldsymbol{A}^T & \boldsymbol{I} \end{pmatrix}, \qquad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{Q}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{F}^{-1} \end{pmatrix}, \tag{2.1}$$

where $\boldsymbol{B}$ is $m \times m$ upper triangular matrix with 1s along its main diagonal, and $\boldsymbol{A}$ is $m \times n$ matrix. The block-diagonal matrix $\boldsymbol{D}$ comprises the inverse of the error covariance matrix $\boldsymbol{Q}$ of $\boldsymbol{\zeta}$ and $\boldsymbol{F}$ of $\boldsymbol{\xi}$, where $\boldsymbol{Q}$ itself is a diagonal matrix. For this purpose we use the block Cholesky decomposition with block sizes $1, \ldots, 1, n$ with number[1] m of 1s.

Wold's algorithm also provides the outer relation matrices $\boldsymbol{C}$ and $\boldsymbol{G}$. In this way, we can estimate the parameter matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{G}$ based on a training sample and on adjacency matrices that specify which latent variable is related to which observable one, both among the exogenous and endogenous variables. This is the point where the expert can intervene the system. For a detailed description, see Section 4.

In the heart of the estimation and the forthcoming filtering there is the simultaneous usage of OLS (ordinary least squares) regression, tracing back to the Gauss normal equations. We give a short summary of that.

Now we concentrate on linear estimates in Hilbert spaces that are the best whenever the underlying distribution is multivariate Gaussian, but is also applicable to second order processes.

Lemma 2.1. *Let $\mathbf{Y} \in \mathbb{R}^p$ and $\mathbf{X} \in \mathbb{R}^q$ be random vectors on a joint probability space with existing second moments and zero expectation. Then*
$\mathbb{E}\|\mathbf{Y} - \boldsymbol{A}^T\mathbf{X}\|^2$ is minimized with

$$\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T], \tag{2.2}$$

[1]Number of endogenous latent variables in the model.

where $\boldsymbol{A}$ is a $q \times p$ matrix and we use generalized inverse $^-$ if the covariance matrix $\mathbb{E}\mathbf{X}\mathbf{X}^T$ of $\mathbf{X}$ is singular. If it is positive definite, then we get a unique estimate for $\boldsymbol{A}$ with the unique inverse matrix $[\mathbb{E}\mathbf{X}\mathbf{X}^T]^{-1}$.

Note that the notation $^-$ applies to any (not necessarily unique) generalized inverse, whereas $^+$ will be used for the uniquely defined Moore–Penrose generalized inverse, see [6].

Proof. Observe that minimizing

$$\mathbb{E}\|\mathbf{Y} - \boldsymbol{A}^T\mathbf{X}\|^2 = \sum_{i=1}^{p}(Y^i - \mathbf{a}_i^T\mathbf{X})^2$$

with respect to $\boldsymbol{A} = (\mathbf{a}_1 \dots \mathbf{a}_p)$ falls apart into the following p minimization tasks, with respect to the q-dimensional column vectors of $\boldsymbol{A}$:

$$\min_{\mathbf{a}_i}(Y^i - \mathbf{a}_i^T\mathbf{X})^2, \quad i = 1, \dots, p.$$

The solution (e.g., with the help of differentiation) gives the well known *Gauss normal equations* from the classical theory of multivariate regression:

$$[\mathbb{E}\mathbf{X}\mathbf{X}^T]\mathbf{a}_i = [\mathbb{E}\mathbf{X}Y_i], \quad i = 1, \dots, p.$$

Since this system of linear equations is consistent (the vector $\mathbb{E}\mathbf{X}Y_i$ is in the column space of $\mathbb{E}\mathbf{X}\mathbf{X}^T$), it always has a solution in the general form:

$$\mathbf{a}_i = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}Y_i], \quad i = 1, \dots, p.$$

Therefore the matrix $\boldsymbol{A}$ giving the optimum is

$$\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T],$$

that is unique only if $\mathbb{E}\mathbf{X}\mathbf{X}^T$ is invertible (positive definite), otherwise (if $\mathbb{E}\mathbf{X}\mathbf{X}^T$ is singular, positive semidefinite) infinitely many versions of the generalized inverse give infinitely many convenient $\boldsymbol{A}$s. However, these always provide the same optimal linear prediction (projection) for $\mathbf{Y}$ as follows:

$$\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y} = \hat{\mathbf{Y}} = \begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_p \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{X} \\ \mathbf{a}_2^T\mathbf{X} \\ \vdots \\ \mathbf{a}_p^T\mathbf{X} \end{bmatrix} = \boldsymbol{A}^T\mathbf{X},$$

where $\mathrm{Proj}_{H(\mathbf{X})}$ denotes the projection onto the Hilbert space spanned by the linear combinations of the components of $\mathbf{X}$ (the expectations are zeros and the inner product is the covariance). □

Lemma 2.2. *Let $\mathbf{Y} \in \mathbb{R}^p$ and $\mathbf{X} \in \mathbb{R}^q$ be random vectors on a joint probability space with existing second moments and zero expectation, and let $\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}$ denotes the best linear prediction of $\mathbf{Y}$ based on $\mathbf{X}$, as before. Then with any $p \times p$ matrix $\boldsymbol{\Phi}$,*

$$\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}.$$

Proof. We saw that $\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y} = \boldsymbol{A}^T\mathbf{X}$, where by (2.2), $\boldsymbol{A} = [\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}\mathbf{Y}^T]$, and we use the generalized inverse $^-$ if the covariance matrix $\mathbb{E}\mathbf{X}\mathbf{X}^T$ of $\mathbf{X}$ is singular. Then

$$\begin{aligned}\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) &= \{[\mathbb{E}\mathbf{X}\mathbf{X}^T]^-[\mathbb{E}\mathbf{X}(\boldsymbol{\Phi}\mathbf{Y})^T]\}^T\mathbf{X} = [\mathbb{E}(\boldsymbol{\Phi}\mathbf{Y}\mathbf{X}^T)][\mathbb{E}\mathbf{X}\mathbf{X}^T]^-\mathbf{X}\\ &= \boldsymbol{\Phi}[\mathbb{E}(\mathbf{Y}\mathbf{X}^T)][\mathbb{E}\mathbf{X}\mathbf{X}^T]^-\mathbf{X} = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}\mathbf{Y}.\end{aligned}$$

□

The above lemma shows that this projection is linear in $\mathbf{Y}$ and it commutes with $\boldsymbol{\Phi}$. In the Gaussian case, obviously, we have that

$$\mathrm{Proj}_{H(\mathbf{X})}(\boldsymbol{\Phi}\mathbf{Y}) = \mathbb{E}(\boldsymbol{\Phi}\mathbf{Y}\,|\,\mathbf{X}) = \boldsymbol{\Phi}\mathbb{E}(\mathbf{Y}\,|\,\mathbf{X}) = \boldsymbol{\Phi}\mathrm{Proj}_{H(\mathbf{X})}(\mathbf{Y})$$

by the properties of the conditional expectation.

The above setup is used for simultaneous (in other words, multiple response) regressions when we regress the components of a random vector (target) with all the components of the predictors.

3. The linear dynamical system for the prediction

Discrete time observations $\mathbf{X}_t$, $\mathbf{Y}_t$ arrive, whereas $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}_t$ are latent state variables corresponding to them. Starting at time 0, for $t = 1, 2, \ldots$, the estimate of $\hat{\boldsymbol{\eta}}_t$ is found, while observing $\mathbf{X}_1, \ldots, \mathbf{X}_t$. Actually, to find $\hat{\boldsymbol{\eta}}_t$, we only need the estimate $\hat{\boldsymbol{\xi}}_t$ and the last observation $\mathbf{X}_t$. Then to find $\hat{\boldsymbol{\xi}}_{t+1}$, the preceding estimate $\hat{\boldsymbol{\eta}}_t$ and the last observation $\mathbf{Y}_t$ are needed. In this way, a recursion is given via the propagation of the error covariance matrices. During the calculations, we use the linearity of the state equations and the predictions, for which we confine ourselves to the second moments of the underlying distributions (second order processes).

The linear dynamical system is

$$\begin{aligned}\boldsymbol{B}\boldsymbol{\eta}_t &= \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t\\ \boldsymbol{U}\boldsymbol{\xi}_{t+1} &= \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t\\ \mathbf{X}_t &= \boldsymbol{C}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t\\ \mathbf{Y}_t &= \boldsymbol{G}\boldsymbol{\eta}_t + \boldsymbol{\delta}_t,\end{aligned} \tag{3.1}$$

where $\boldsymbol{A}$ is $m \times n$, $\boldsymbol{B}$ is $m \times m$, $\boldsymbol{V}$ is $n \times m$, $\boldsymbol{U}$ is $n \times n$, $\boldsymbol{C}$ is $q \times n$, and $\boldsymbol{G}$ is $p \times m$ specified matrix; $\boldsymbol{B}$ and $\boldsymbol{U}$ are non-singular (in recursive models they are upper triangular with 1s along their main diagonals). Further, $\boldsymbol{\zeta}_t$ is an orthogonal process with $\mathbb{E}\boldsymbol{\zeta}_t\boldsymbol{\zeta}_s^T = \delta_{st}\boldsymbol{Q}$ with diagonal covariance matrix $\boldsymbol{Q}$; $\boldsymbol{\gamma}_t$ is an orthogonal process with $\mathbb{E}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_s^T = \delta_{st}\boldsymbol{R}$ with diagonal covariance matrix $\boldsymbol{R}$; $\mathbb{E}\boldsymbol{\xi}_s^T\boldsymbol{\zeta}_t = \mathbf{0}$ and $\mathbb{E}\boldsymbol{\eta}_s^T\boldsymbol{\gamma}_t = \mathbf{0}$ for $s \le t$; $\boldsymbol{\varepsilon}_t$ is independent of $\boldsymbol{\xi}_t$, $\boldsymbol{\delta}_t$ is independent of $\boldsymbol{\eta}_t$, they are also independent of each other and of $\boldsymbol{\zeta}_t$ and $\boldsymbol{\gamma}_t$. For simplicity, we assume that all the expectations are zeros.

The $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{U}, \boldsymbol{V}$ matrices are estimated from a training sample. Actually, the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ together with $\boldsymbol{Q}$ and $\boldsymbol{F}$ come from the block Cholesky decomposition (2.1), based on the product-moments of the estimated latent scores of the pairs $\boldsymbol{\xi}_s, \boldsymbol{\eta}_s$, where $s < 0$ is integer from the past (training sample). Likewise, the matrices

$\boldsymbol{U}$ and $\boldsymbol{V}$ together with $\boldsymbol{R}$ and $\boldsymbol{F}^*$ come from the block Cholesky decomposition (3.2) below, based on the product-moments of the estimated latent scores of the shifted pairs $\boldsymbol{\eta}_s, \boldsymbol{\xi}_{s+1}$ $(s < 0)$. The inverse of this matrix is decomposed as $\boldsymbol{L}^*\boldsymbol{D}^*\boldsymbol{L}^{*T}$ with $\boldsymbol{L}^*$ and $\boldsymbol{D}^*$ having the form

$$\boldsymbol{L}^* = \begin{pmatrix} \boldsymbol{U}^T & \boldsymbol{O} \\ -\boldsymbol{V}^T & \boldsymbol{I} \end{pmatrix}, \qquad \boldsymbol{D}^* = \begin{pmatrix} \boldsymbol{R}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{F}^{*-1} \end{pmatrix}, \tag{3.2}$$

where recall that, in the recursive model, $\boldsymbol{U}$ is $n \times n$ upper triangular matrix with 1s along its main diagonal, and $\boldsymbol{V}$ is $n \times m$. The block-diagonal matrix $\boldsymbol{D}^*$ comprises the inverse of the error covariance matrix $\boldsymbol{R}$ of $\boldsymbol{\gamma}$ and $\boldsymbol{F}^*$ of $\boldsymbol{\eta}$, where $\boldsymbol{R}$ itself is a diagonal matrix. For this purpose we use the block Cholesky decomposition with block sizes $1, \dots, 1, m$ with number[2] n of 1s.

The matrices $\boldsymbol{C}$ ad $\boldsymbol{G}$ are estimated by the PLS algorithm of H. Wold [8]. The details are given in Section 4.

Now a recursion is introduced for the following problem: starting the observations at time 0 in the test sample, we want to estimate $\boldsymbol{\eta}_t$ based on $\mathbf{X}_t$, and $\boldsymbol{\xi}_{t+1}$ based on $\mathbf{Y}_t$ component-wise, with minimum mean square error. Former observations also play role, but only through the last one and through the propagation of the error covariance matrices. Here $\mathbf{X}_0$ and $\mathbf{Y}_0$ can be taken from the training sample.

3.1. First stage: $\mathbf{X}_t \to \hat{\boldsymbol{\eta}}_t$

For $t \geq 1$, let $H_{t-1}(\mathbf{X}) = \mathrm{Span}\,(\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$ consists of the linear combinations of all the components of $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ over a common probability space. They are also in a Hilbert space (L_2 space) with the covariance as inner product. We denote the optimal prediction of $\boldsymbol{\xi}_t$ based on $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ by $\hat{\boldsymbol{\xi}}_t$.

If $\mathbf{X}_0, \dots, \mathbf{X}_{t-1}$ are observed, i.e., $H_{t-1}(\mathbf{X})$ is known, then the newly observed (measured) $\mathbf{X}_t$ can be orthogonally decomposed as

$$\mathbf{X}_t = \mathrm{Proj}_{H_{t-1}(\mathbf{X})}\mathbf{X}_t + \tilde{\mathbf{X}}_t = \overline{\mathbf{X}}_t + \tilde{\mathbf{X}}_t, \tag{3.3}$$

where the orthogonal component $\tilde{\mathbf{X}}_t \in I_t(\mathbf{X})$, and $I_t(\mathbf{X})$ is the so-called innovation subspace (actually, the components of $\tilde{\mathbf{X}}_t$ generate $I_t(\mathbf{X})$). Assume that $I_t(\mathbf{X})$ is not the sole $\mathbf{0}$ vector, otherwise observing $\mathbf{X}_t$ does not give any additional information to $H_{t-1}(\mathbf{X})$. If $\{\mathbf{X}_t\}$ is weakly stationary, it means that the process is *regular*.

Equation (3.3) implies the decomposition of the corresponding subspaces like

$$H_t(\mathbf{X}) = H_{t-1}(\mathbf{X}) \oplus I_t(\mathbf{X}), \tag{3.4}$$

that is the analogue of multidimensional Wold decomposition when we make one-step ahead prediction based on finitely many past values. (The Wold decomposition applies to the stationary and infinite past case. Indeed, when $t \to \infty$, i.e., going to the future, we approach this situation in the stationary case).

[2]Number of exogenous latent variables in the model.

Assume that we have already found $\hat{\boldsymbol{\xi}}_t$. We shall give a recursion to find $\hat{\boldsymbol{\eta}}_t$ by using the new value of $\mathbf{X}_t$. In view of Equation (3.4), we proceed as follows:

$$\begin{aligned}\boldsymbol{B}\hat{\boldsymbol{\eta}}_t &= \mathrm{Proj}_{H_t(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t) = \mathrm{Proj}_{H_{t-1}(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t) + \mathrm{Proj}_{I_t(\mathbf{X})}(\boldsymbol{B}\boldsymbol{\eta}_t) \\ &= \boldsymbol{A}\mathrm{Proj}_{H_{t-1}(\mathbf{X})}\boldsymbol{\xi}_t + \mathrm{Proj}_{H_{t-1}(\mathbf{X})}\boldsymbol{\zeta}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t \\ &= \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t,\end{aligned} \tag{3.5}$$

where we utilized that $\boldsymbol{\zeta}_t \perp H_{t-1}(\mathbf{X})$, Lemma 2.1 and the first state equation of (3.1). We refer to the linearity of the projection, see Lemma 2.2. Since $\mathrm{Proj}_{I_t(\mathbf{X})}\boldsymbol{B}\boldsymbol{\eta}_t$ is the linear combination of the coordinates of the vector $\tilde{\mathbf{X}}_t \in I_t(\mathbf{X})$, its effect can be written as a matrix $\boldsymbol{K}_t$ multiplied with $\tilde{\mathbf{X}}_t$. This $m \times q$ matrix $\boldsymbol{K}_t$ is called *Kálmán gain matrix* after R. E. Kálmán (in fact, this notation was first used in the paper [4] of Kálmán and Bucy).

To specify the matrix $\boldsymbol{K}_t$, we have to write $\tilde{\mathbf{X}}_t$ in terms of $\hat{\boldsymbol{\xi}}_t$ and $\mathbf{X}_t$. For this purpose, let us project both sides of the first observation equation of (3.1), i.e., of $\mathbf{X}_t = \boldsymbol{C}\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_t$, onto $H_{t-1}(\mathbf{X})$. We get that

$$\overline{\mathbf{X}}_t = \boldsymbol{C}\hat{\boldsymbol{\xi}}_t.$$

Taking the orthogonal decomposition (3.3) of $\mathbf{X}_t$ into consideration yields that

$$\tilde{\mathbf{X}}_t = \mathbf{X}_t - \overline{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t. \tag{3.6}$$

We substitute this into the last line of Equation (3.5) and obtain that

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t = (\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t.$$

With the notation

$$\boldsymbol{A}_t^* = \boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C} \tag{3.7}$$

for the updated transition matrix, we get the new linear dynamics:

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}_t^*\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t. \tag{3.8}$$

We also have the alternative expression

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\tilde{\mathbf{X}}_t = \boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t). \tag{3.9}$$

The estimation error is also governed by the linear dynamical system. This error term has two alternative forms. Using (3.8), the one is

$$\begin{aligned}\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t &= \boldsymbol{B}\boldsymbol{\eta}_t - \boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t - \boldsymbol{A}^*\hat{\boldsymbol{\xi}}_t - \boldsymbol{K}_t\boldsymbol{C}\boldsymbol{\xi}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t \\ &= \boldsymbol{A}_t^*(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_t) + \boldsymbol{\zeta}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t = \boldsymbol{A}_t^*\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t\boldsymbol{\varepsilon}_t.\end{aligned}$$

Then, using (3.9), the other is

$$\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t = \boldsymbol{A}\boldsymbol{\xi}_t + \boldsymbol{\zeta}_t - \boldsymbol{A}\hat{\boldsymbol{\xi}}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t) = \boldsymbol{A}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t).$$

From here, we get the following recursion for the covariance matrix

$$\boldsymbol{P}_t = \mathbb{E}\tilde{\boldsymbol{\xi}}_t\tilde{\boldsymbol{\xi}}_t^T \tag{3.10}$$

of the optimal error (of predicting $\boldsymbol{\xi}_t$) and so, of $\boldsymbol{K}_t$:

$$\begin{aligned}\boldsymbol{B}[\mathbb{E}\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T]\boldsymbol{B}^T &= \mathbb{E}[\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t][\boldsymbol{B}\tilde{\boldsymbol{\eta}}_t]^T \\ &= \mathbb{E}[\boldsymbol{A}^*\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t][\boldsymbol{A}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\zeta}_t - \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t)]^T \\ &= \boldsymbol{A}_t^*\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q},\end{aligned} \tag{3.11}$$

where recall that $\boldsymbol{Q} = \mathbb{E}\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t^T$, obtainable by (2.1). We used that $\boldsymbol{\zeta}_t$ is uncorrelated with $\boldsymbol{\xi}_t$ and, therefore, with $\tilde{\boldsymbol{\xi}}_t$ too. We also used that $\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t$ is in $\boldsymbol{I}_t(\mathbf{X})$, and $\boldsymbol{\zeta}_t$ is uncorrelated with $\boldsymbol{\varepsilon}_t$.

It remains to find an explicit formula for $\boldsymbol{K}_t$, and thus, also for $\boldsymbol{A}_t^*$. Recall that $\boldsymbol{K}_t$ is the matrix of the linear operation $\mathrm{Proj}_{I_t(\mathbf{X})}\boldsymbol{B}\boldsymbol{\eta}_t$, therefore by the projection principle (see Lemma 2.1):

$$\boldsymbol{K}_t = [\mathbb{E}\boldsymbol{B}\boldsymbol{\eta}_t\tilde{\mathbf{X}}_t^T][\mathbb{E}(\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t^T]^+,$$

where $^+$ denotes the Moore–Penrose generalized inverse (we use regular inverse if the underlying matrix is invertible).

Now we calculate the matrices in brackets. By the third equation of (3.1), that extends to $\tilde{\mathbf{X}}_t = \boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t$ and to their predictions, we get that

$$\mathbb{E}\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}_t^T = \mathbb{E}(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t)(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t + \boldsymbol{\varepsilon}_t)^T = \boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E},$$

where $\boldsymbol{E} = \mathbb{E}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^T$. $\boldsymbol{E}$ is obtainable by (2.1) in the following way:

$$\mathbb{E}\mathbf{X}_t\mathbf{X}_t^T = \boldsymbol{C}(\mathbb{E}\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T)\boldsymbol{C}^T + \boldsymbol{E} = \boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^T + \boldsymbol{E}.$$

So $\boldsymbol{E}$ is the difference between $\mathbb{E}\mathbf{X}_t\mathbf{X}_t^T$ (estimated as $\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}$ from the training sample) and $\boldsymbol{C}\boldsymbol{F}\boldsymbol{C}^T$, where $\boldsymbol{F}$ is the inverse of the second diagonal block of $\boldsymbol{D}$ in (2.1).

By the first and third equation of (3.1) and the orthogonality of $\hat{\boldsymbol{\xi}}_t$ and $\tilde{\boldsymbol{\xi}}_t$ we get that

$$\mathbb{E}(\boldsymbol{B}\boldsymbol{\eta}_t\tilde{\mathbf{X}}_t^T) = \boldsymbol{A}\mathbb{E}(\boldsymbol{\xi}_t\tilde{\mathbf{X}}_t^T) = \boldsymbol{A}\mathbb{E}[(\hat{\boldsymbol{\xi}}_t + \tilde{\boldsymbol{\xi}}_t)(\boldsymbol{C}\tilde{\boldsymbol{\xi}}_t)^T] = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T. \tag{3.12}$$

Therefore,

$$\boldsymbol{K}_t = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+ \tag{3.13}$$

with the Moore–Penrose inverse.

With this matrix $\boldsymbol{K}_t$ of Equation (3.13) and using Equation (3.11), we are able to write the error covariance matrix in the form of a symmetric matrix:

$$\begin{aligned}\boldsymbol{B}[\mathbb{E}\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T]\boldsymbol{B}^T &= \boldsymbol{A}^*\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} = (\boldsymbol{A} - \boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \left(\boldsymbol{A} - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\right)\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \boldsymbol{A}\left(\boldsymbol{I} - \boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\right)\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q} \\ &= \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q},\end{aligned}$$

so

$$\boldsymbol{B}\boldsymbol{P}_t^*\boldsymbol{B}^T = \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}, \tag{3.14}$$

where $\boldsymbol{P}_t^* = \mathbb{E}(\tilde{\boldsymbol{\eta}}_t\tilde{\boldsymbol{\eta}}_t^T)$ is the covariance matrix of the error when predicting $\boldsymbol{\eta}_t$. In the next stage, we use it to find $\boldsymbol{P}_{t+1}$.

3.2. Second stage: $\mathbf{Y}_t \to \hat{\boldsymbol{\xi}}_{t+1}$

For $t \geq 1$, let $H_{t-1}(\mathbf{Y}) = \mathrm{Span}\,(\mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1})$ consists of the linear combinations of all the components of $\mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}$ over a common probability space. We denote the optimal prediction of $\boldsymbol{\eta}_t$ based on $\mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}$ by $\check{\boldsymbol{\eta}}_t$.

If $\mathbf{Y}_0, \ldots, \mathbf{Y}_{t-1}$ are observed, i.e., $H_{t-1}(\mathbf{Y})$ is known, then the newly observed (measured) $\mathbf{Y}_t$ can be orthogonally decomposed as

$$\mathbf{Y}_t = \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\mathbf{Y}_t + \tilde{\mathbf{Y}}_t = \overline{\mathbf{Y}}_t + \tilde{\mathbf{Y}}_t, \tag{3.15}$$

where the orthogonal component $\tilde{\mathbf{Y}}_t \in I_t(\mathbf{Y})$, and $I_t(\mathbf{Y})$ is the innovation subspace (actually, the components of $\tilde{\mathbf{Y}}_t$ generate $I_t(\mathbf{Y})$). Assume that $I_t(\mathbf{Y})$ is not the sole $\mathbf{0}$ vector, otherwise observing $\mathbf{Y}_t$ does not give any additional information to $H_{t-1}(\mathbf{Y})$.

Equation (3.15) implies the decomposition of the corresponding subspaces like

$$H_t(\mathbf{Y}) = H_{t-1}(\mathbf{Y}) \oplus I_t(\mathbf{Y}). \tag{3.16}$$

Assume that we have already found $\check{\boldsymbol{\eta}}_t$. We shall give a recursion to find $\check{\boldsymbol{\xi}}_{t+1}$ by using the new value of $\mathbf{Y}_t$. In view of Equation (3.16):

$$\begin{aligned}\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} &= \mathrm{Proj}_{H_t(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) = \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) + \mathrm{Proj}_{I_t(\mathbf{Y})}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}) \\ &= \boldsymbol{V}\mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\boldsymbol{\eta}_t + \mathrm{Proj}_{H_{t-1}(\mathbf{Y})}\boldsymbol{\gamma}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t \\ &= \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t,\end{aligned} \tag{3.17}$$

where we utilized that $\boldsymbol{\gamma}_t \perp H_{t-1}(\mathbf{Y})$, Lemma 2.1 and the second state equation of (3.1). Furthermore, we refer to the linearity of the projection, see Lemma 2.2. Since $\mathrm{Proj}_{I_t(\mathbf{Y})}\boldsymbol{U}\boldsymbol{\xi}_{t+1}$ is the linear combination of the coordinates of the vector $\tilde{\mathbf{Y}}_t \in I_t(\mathbf{Y})$, its effect can be written as a matrix $\boldsymbol{M}_t$ multiplied with $\tilde{\mathbf{Y}}_t$. This $n \times p$ matrix $\boldsymbol{M}_t$ is another gain matrix.

To specify the matrix $\boldsymbol{M}_t$, we have to write $\tilde{\mathbf{Y}}_t$ in terms of $\check{\boldsymbol{\eta}}_t$ and $\mathbf{Y}_t$. For this purpose, let us project both sides of the second observation equation of (3.1), i.e., of $\mathbf{Y}_t = \boldsymbol{G}\boldsymbol{\eta}_t + \boldsymbol{\delta}_t$, onto $H_{t-1}(\mathbf{Y})$. We get that

$$\overline{\mathbf{Y}}_t = \boldsymbol{G}\check{\boldsymbol{\eta}}_t.$$

Taking the orthogonal decomposition (3.15) of $\mathbf{Y}_t$ into consideration yields that

$$\tilde{\mathbf{Y}}_t = \mathbf{Y}_t - \overline{\mathbf{Y}}_t = \mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t. \tag{3.18}$$

We substitute this into the last line of Equation (3.17) and obtain that

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t = (\boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G})\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t.$$

With the notation

$$\boldsymbol{V}_t^* = \boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G} \tag{3.19}$$

for the updated transition matrix, we get the new linear dynamics:

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}_t^*\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t. \tag{3.20}$$

We also have the alternative expression

$$\boldsymbol{U}\check{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\tilde{\mathbf{Y}}_t = \boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t). \tag{3.21}$$

The estimation error is also governed by the linear dynamical system. This error term has two alternative forms. Using (3.20), the one is

$$\begin{aligned}\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1} &= \boldsymbol{U}\boldsymbol{\xi}_{t+1} - \boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t - \boldsymbol{V}_t^*\breve{\boldsymbol{\eta}}_t - \boldsymbol{M}_t\boldsymbol{G}\boldsymbol{\eta}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t \\ &= \boldsymbol{V}_t^*(\boldsymbol{\eta}_t - \breve{\boldsymbol{\eta}}_t) + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t = \boldsymbol{V}_t^*\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t.\end{aligned}$$

Then, using (3.21), the other is

$$\boldsymbol{U}\breve{\breve{\boldsymbol{\xi}}}_{t+1} = \boldsymbol{V}\boldsymbol{\eta}_t + \boldsymbol{\gamma}_t - \boldsymbol{V}\breve{\boldsymbol{\eta}}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\breve{\boldsymbol{\eta}}_t) = \boldsymbol{V}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{M}_t\breve{\boldsymbol{\eta}}_t).$$

From here, we get the following recursion for the covariance matrix

$$\boldsymbol{P}_t^* = \mathbb{E}\breve{\boldsymbol{\eta}}_t\breve{\boldsymbol{\eta}}_t^T \tag{3.22}$$

of the optimal error (of predicting $\boldsymbol{\eta}_t$) and so, of $\boldsymbol{M}_t$:

$$\begin{aligned}\boldsymbol{U}[\mathbb{E}\breve{\boldsymbol{\xi}}_{t+1}\breve{\boldsymbol{\xi}}_{t+1}^T]\boldsymbol{U}^T &= \mathbb{E}[\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1}][\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1}]^T \\ &= \mathbb{E}[\boldsymbol{V}^*\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t\boldsymbol{\delta}_t][\boldsymbol{V}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\gamma}_t - \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\breve{\boldsymbol{\eta}}_t)]^T \\ &= \boldsymbol{V}_t^*\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R},\end{aligned} \tag{3.23}$$

where recall that $\boldsymbol{R} = \mathbb{E}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^T$, obtainable by (3.2), and we used that $\boldsymbol{\gamma}_t$ is uncorrelated with $\boldsymbol{\eta}_t$ and, therefore, with $\breve{\boldsymbol{\eta}}_t$ too; further, $\boldsymbol{V}_t^* = \boldsymbol{V} - \boldsymbol{M}_t\boldsymbol{G}$. We also used that $\mathbf{Y}_t - \boldsymbol{G}\breve{\boldsymbol{\eta}}_t$ is in $\boldsymbol{I}_t(\mathbf{Y})$, and that $\boldsymbol{\gamma}_t$ is uncorrelated with $\boldsymbol{\delta}_t$.

Now an explicit formula is found for $\boldsymbol{M}_t$, and thus, also for $\boldsymbol{V}_t^*$. Recall that $\boldsymbol{M}_t$ is the matrix of the linear operation $\mathrm{Proj}_{I_t(\mathbf{Y})}\boldsymbol{U}\boldsymbol{\xi}_{t+1}$, therefore by the projection principle (see Lemma 2.1):

$$\boldsymbol{M}_t = [\mathbb{E}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}\tilde{\mathbf{Y}}_t^T)][\mathbb{E}(\tilde{\mathbf{Y}}_t\tilde{\mathbf{Y}}_t^T]^+,$$

where $^+$ denotes the Moore–Penrose generalized inverse (we use regular inverse if the underlying matrix is invertible). We calculate the matrices in brackets. By the last equation of (3.1), that extends to $\tilde{\mathbf{Y}}_t = \boldsymbol{G}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t$ and to their predictions, we get that

$$\mathbb{E}(\tilde{\mathbf{Y}}_t\tilde{\mathbf{Y}}_t^T) = \mathbb{E}[(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t)(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t + \boldsymbol{\delta}_t)^T = \boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta},$$

where $\boldsymbol{\Delta} = \mathbb{E}\boldsymbol{\delta}_t\boldsymbol{\delta}_t^T$. $\boldsymbol{\Delta}$ is obtainable by (3.2) in the following way:

$$\mathbb{E}\mathbf{Y}_t\mathbf{Y}_t^T = \boldsymbol{G}(\mathbb{E}\boldsymbol{\eta}_t\boldsymbol{\eta}_t^T)\boldsymbol{G}^T + \Delta = \boldsymbol{G}\boldsymbol{F}^*\boldsymbol{G}^T + \boldsymbol{\Delta}.$$

So $\boldsymbol{\Delta}$ is the difference between $\mathbb{E}\mathbf{Y}_t\mathbf{Y}_t^T$ (estimated as $\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}$ from the training sample) and $\boldsymbol{G}\boldsymbol{F}^*\boldsymbol{G}^T$, where $\boldsymbol{F}^*$ is the inverse of the second diagonal block of $\boldsymbol{D}^*$ in (3.2).

By the second and fourth equation of (3.1) and the orthogonality of $\breve{\boldsymbol{\eta}}_t$ and $\breve{\boldsymbol{\eta}}_t$ we get that

$$\mathbb{E}(\boldsymbol{U}\boldsymbol{\xi}_{t+1}\tilde{\mathbf{Y}}_t^T) = \boldsymbol{V}\mathbb{E}(\boldsymbol{\eta}_t\tilde{\mathbf{Y}}_t^T) = \boldsymbol{V}\mathbb{E}[(\breve{\boldsymbol{\eta}}_t + \breve{\boldsymbol{\eta}}_t)(\boldsymbol{G}\breve{\boldsymbol{\eta}}_t)^T] = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T. \tag{3.24}$$

Therefore,

$$\boldsymbol{M}_t = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+ \tag{3.25}$$

with the Moore–Penrose inverse.

With this matrix $\boldsymbol{M}_t$ of Equation (3.25) and using Equation (3.23), we are able to write the error covariance matrix in the form of a symmetric matrix:

$$\begin{aligned}\boldsymbol{U}[\mathbb{E}\breve{\boldsymbol{\xi}}_{t+1}\breve{\boldsymbol{\xi}}_{t+1}^T]\boldsymbol{U}^T &= \boldsymbol{V}_t^*\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{\Delta} = (\boldsymbol{V}-\boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= (\boldsymbol{V}-\boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T+\boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= \boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P}_t^*\boldsymbol{G}^T)[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T+\boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}\\ &= \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T+\boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R},\end{aligned}$$

so

$$\boldsymbol{U}\boldsymbol{P}_{t+1}\boldsymbol{U}^T = \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T+\boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}, \tag{3.26}$$

where we assumed that the error covariance matrix of $\tilde{\boldsymbol{\xi}}_t$ and $\breve{\boldsymbol{\xi}}_t$, akin to that of $\tilde{\boldsymbol{\eta}}_t$ and $\breve{\boldsymbol{\eta}}_t$ is the same. This fact gives rise to a recursion by connecting (3.14) and (3.26).

Finally, with (3.9) and (3.21) we are able to recursively estimate the latent state variables. During the $\boldsymbol{P}_1 \to \boldsymbol{P}_1^* \to \boldsymbol{P}_2 \to \boldsymbol{P}_2^* \ldots$ recursion, from $\boldsymbol{P}_t$, we find $\boldsymbol{K}_t$ by (3.13) and $\hat{\boldsymbol{\eta}}_t$ by (3.9). Then, from $\boldsymbol{P}_t^*$, we find $\boldsymbol{M}_t$ by (3.25) and $\breve{\boldsymbol{\xi}}_{t+1}$ by (3.21).

As for the relation between $\hat{\boldsymbol{\xi}}_t$ and $\breve{\boldsymbol{\xi}}_t$, akin to that between $\hat{\boldsymbol{\eta}}_t$ and $\breve{\boldsymbol{\eta}}_t$, we can estimate their cross-covariance matrices from the training sample, and then, linearly predict $\breve{\boldsymbol{\eta}}_t$ with $\hat{\boldsymbol{\eta}}_t$ and linearly predict $\breve{\boldsymbol{\xi}}_t$ with $\hat{\boldsymbol{\xi}}_t$ by Lemma 2.1 as follows:

$$\breve{\boldsymbol{\eta}}_t = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\mathbf{Y}}\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\mathbf{Y}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\mathbf{X}}\hat{\boldsymbol{\Sigma}}_{\mathbf{X}\mathbf{X}}^+\boldsymbol{\Sigma}_{\mathbf{X}\boldsymbol{\eta}}[\mathbb{E}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^T]^+\hat{\boldsymbol{\eta}}_t$$

and

$$\hat{\boldsymbol{\xi}}_{t+1} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}_{\mathbf{X}\mathbf{X}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{X}\mathbf{Y}}\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\mathbf{Y}}^+\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\boldsymbol{\xi}}[\mathbb{E}\breve{\boldsymbol{\xi}}_{t+1}\breve{\boldsymbol{\xi}}_{t+1}^T]^+\breve{\boldsymbol{\xi}}_{t+1}.$$

Here, from Equation (3.11), we conclude that

$$\mathbb{E}\hat{\boldsymbol{\eta}}_t\hat{\boldsymbol{\eta}}_t^T = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\boldsymbol{\eta}} - \boldsymbol{B}^{-1}[(\boldsymbol{A}-\boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}](\boldsymbol{B}^{-1})^T.$$

Likewise, from Equation (3.23), we conclude that

$$\mathbb{E}\breve{\boldsymbol{\xi}}_t\breve{\boldsymbol{\xi}}_t^T = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{U}^{-1}[(\boldsymbol{V}-\boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}_t^T + \boldsymbol{R}](\boldsymbol{U}^{-1})^T.$$

3.3. The main result

We assume that the system was at rest until time 0. The parameter matrices are estimated from the past, whereas newer and newer estimates for the latent variables are given, as observations arrive at time t ($t = 1, 2, \ldots$ up to the end of the experimental time T). Thus, we can summarize the results in the subsequent theorem. It is important that in the derivation of the formulas we used the best linear prediction theory of Hilbert spaces.

Theorem 3.1. *In the linear dynamical system (3.1), the optimal estimate $\hat{\boldsymbol{\eta}}_t$ of $\boldsymbol{\eta}_t$ and $\breve{\boldsymbol{\xi}}_{t+1}$ of $\boldsymbol{\xi}_{t+1}$ given $\mathbf{X}_1, \ldots, \mathbf{X}_t$ and $\mathbf{Y}_1, \ldots, \mathbf{Y}_t$ is generated by the new linear dynamical system*

$$\boldsymbol{B}\hat{\boldsymbol{\eta}}_t = \boldsymbol{A}_t^*\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t\mathbf{X}_t$$

and

$$\boldsymbol{U}\breve{\boldsymbol{\xi}}_{t+1} = \boldsymbol{V}_t^*\breve{\boldsymbol{\eta}}_t + \boldsymbol{M}_t\mathbf{Y}_t.$$

The expected quadratic losses are $tr\boldsymbol{P}_t^*$ and $tr\boldsymbol{P}_{t+1}$, where $\boldsymbol{P}_t^*$ and $\boldsymbol{P}_{t+1}$ are the propagated covariance matrices of the estimation errors. The minimizing matrices and the

one-step ahead predictions $\hat{\boldsymbol{\eta}}_t$ *and* $\hat{\boldsymbol{\xi}}_{t+1}$ *together with the error covariance matrices* $\boldsymbol{P}_t^*$ *and* $\boldsymbol{P}_{t+1}$ *are uniquely determined by the initial conditions*

$$\hat{\boldsymbol{\xi}}_1 = \mathrm{Proj}_{\mathbf{X}_0}\boldsymbol{\xi}_1, \quad \tilde{\boldsymbol{\xi}}_1 = \boldsymbol{\xi}_1 - \hat{\boldsymbol{\xi}}_1, \quad \boldsymbol{P}_1 = \mathbb{E}\tilde{\boldsymbol{\xi}}_1\tilde{\boldsymbol{\xi}}_1^T$$

and the recursions for $t = 1, 2, \ldots$ *as follows.*

$$\begin{aligned}
\boldsymbol{K}_t &= \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\\
\hat{\boldsymbol{\eta}}_t &= \boldsymbol{B}^{-1}[\boldsymbol{A}\hat{\boldsymbol{\xi}}_t + \boldsymbol{K}_t(\mathbf{X}_t - \boldsymbol{C}\hat{\boldsymbol{\xi}}_t)]\\
\hat{\mathbf{Y}}_t &= \boldsymbol{G}\hat{\boldsymbol{\eta}}_t\\
\check{\boldsymbol{\eta}}_t &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{YY}}\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}\boldsymbol{\eta}}\left[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\eta}\boldsymbol{\eta}} - \boldsymbol{B}^{-1}((\boldsymbol{A}-\boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q})(\boldsymbol{B}^{-1})^T\right]^+\hat{\boldsymbol{\eta}}_t\\
\boldsymbol{P}_t^* &= \boldsymbol{B}^{-1}[\boldsymbol{A}\boldsymbol{P}_t\boldsymbol{A}^T - \boldsymbol{A}\boldsymbol{P}_t\boldsymbol{C}^T[\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{C}^T + \boldsymbol{E}]^+\boldsymbol{C}\boldsymbol{P}_t\boldsymbol{A}^T + \boldsymbol{Q}]\boldsymbol{B}^{-1^T}\\
\boldsymbol{M}_t &= \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\\
\check{\boldsymbol{\xi}}_{t+1} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\check{\boldsymbol{\eta}}_t + \boldsymbol{M}_t(\mathbf{Y}_t - \boldsymbol{G}\check{\boldsymbol{\eta}}_t)]\\
\hat{\boldsymbol{\xi}}_{t+1} &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{Y}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{YY}}\hat{\boldsymbol{\Sigma}}_{\mathbf{YX}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\hat{\boldsymbol{\Sigma}}_{\mathbf{X}\boldsymbol{\xi}}\left[\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\boldsymbol{\xi}} - \boldsymbol{U}^{-1}((\boldsymbol{V}-\boldsymbol{M}_t\boldsymbol{G})\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R})(\boldsymbol{U}^{-1})^T\right]^+\check{\boldsymbol{\xi}}_{t+1}\\
\hat{\mathbf{X}}_{t+1} &= \boldsymbol{C}\hat{\boldsymbol{\xi}}_{t+1}\\
\boldsymbol{P}_{t+1} &= \boldsymbol{U}^{-1}[\boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{V}^T - \boldsymbol{V}\boldsymbol{P}_t^*\boldsymbol{G}^T[\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{G}^T + \boldsymbol{\Delta}]^+\boldsymbol{G}\boldsymbol{P}_t^*\boldsymbol{V}^T + \boldsymbol{R}]\boldsymbol{U}^{-1^T},
\end{aligned}$$

where $^+$ *denotes the Moore–Penrose generalized inverse (usual inverse if the matrix is invertible).*

Note that $\hat{\boldsymbol{\xi}}_1 = \mathrm{Proj}_{\mathbf{X}_0}\boldsymbol{\xi}_1 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\mathbf{X}_0$, by Lemma 2.1, where the last training sample entry can be chosen for $\mathbf{X}_0$. To initialize $\boldsymbol{P}_1$, the whole training sample can be used: if the L learning sample entries are indexed by ℓ, then $\hat{\boldsymbol{\xi}}_\ell = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\xi}\mathbf{X}}\hat{\boldsymbol{\Sigma}}^+_{\mathbf{XX}}\mathbf{X}_\ell$ and $\tilde{\boldsymbol{\xi}}_\ell = \boldsymbol{\xi}_\ell - \hat{\boldsymbol{\xi}}_\ell$, where $\boldsymbol{\xi}_\ell$ is the ℓth case estimate of $\boldsymbol{\xi}$, based on the forthcoming PLS algorithm in Section 4. Finally, the product moment estimate of $\boldsymbol{P}_1$ is $\frac{1}{L}\sum_{\ell=1}^{L}\tilde{\boldsymbol{\xi}}_\ell\tilde{\boldsymbol{\xi}}_\ell^T$ if the variables have zero expectation (otherwise, the sample means should be subtracted).

Note that the matrices $\mathbf{U}, \mathbf{V}$ are also estimated from the learning sample with the shifted product moments, as discussed in the next section.

4. Application

Using data from three Egyptian villages, we applied our proposed algorithm to examine and predict to what extent parental views affect their daughters' thinking on two empowerment issues. Figure 1 visualizes the hypothesized outer and inner relations. The inner model examines the cause-effect structure between the latent variables (LVs): parental views on girls' participation in decision making and girls' mobility (exogenous: P-DM=ξ^1 and P-Mob=ξ^2) and the daughters' views on the same issues (endogenous: G-DM=η^1 and G-Mob=η^2), respectively. The outer model links LVs and observed variables (OVs) together. Mode A is used to construct all LVs in the model. To clarify, ξ^1 is composed of four independent xs (OVs) that measure parental views on girl's responsibility in making decisions related to marriage,

choosing a husband, entering and continuing schooling. ξ^2 is linked to three xs that ask parents whether girls can go alone to places such as market, field, and friends' home. In the same manner, η^1 and η^2 are connected to the same number of OVs, but the dependent ys that reflect the daughters' views on the same indicators. Note, all the OVs are on the same ordinal scale.

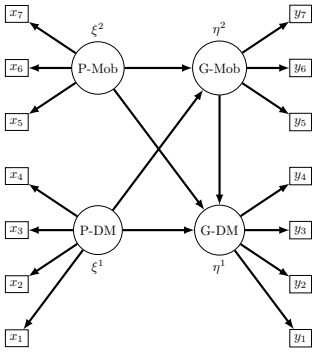

FIGURE 1. Schematic representation of the SEM that considers the effect of parental views on girls' views related to empowerment issues.

A total sample of 349 parents and their daughters are considered. Prior to the analysis, the data were standardized to have zero mean and unit variance. Then, it was divided randomly into a training sample of size 279 and a test sample of the remaining cases. We apply our proposed estimation algorithm on the training sample to obtain the specified parameter matrices ($\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{G}$, $\boldsymbol{Q}$, $\boldsymbol{F}$, $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{R}$, and $\boldsymbol{F}^*$). These matrices are used in the derivation of the proposed prediction recursion algorithm. Specifically, the filtering technique in Theorem 3.1 predicts the future values for the test sample observations that come sequentially.

Recall that our integrated estimation algorithm to obtain the model specified matrices combines the first stage of Wold's PLS technique and the block Cholesky decomposition of Kiiveri et al. The detailed explanation follows:

1. The first stage uses stage I of Wold's PLS algorithm, in which the outer relations and the LV case values are obtained from the training sample. It is an iterative process that consists of the following steps:
   i. Initialize the LV scores for each case as the weighted sum of the observed indicators in the block that correspond to each LV:
   $$\boldsymbol{H} = \boldsymbol{N}\boldsymbol{Z},$$
   where $\boldsymbol{H}$ is the exogenous and endogenous LV scores matrix, $\boldsymbol{N}$ is the training sample data matrix of size 279×14, and $\boldsymbol{Z}$ is the 14×4 adjacency matrix of the measurement model. The entries z_{kj} are ones, if the indicator n_{kj} belongs to the block that defines the corresponding LV; and zeros, otherwise. After each step, the LV scores are standardized.
   ii. Update the obtained matrix $\boldsymbol{H}$ with the inner weights
   $$\tilde{\boldsymbol{H}} = \boldsymbol{H}\boldsymbol{W},$$
   where $\boldsymbol{W}$ is the LV inner weights matrix which is computed for each LV to indicate how strong it is connected to the other LVs in the model. There

are three schemes to obtain these weights, for more details, see [7] and [8]. We used the centroid scheme.

iii. Use the obtained LV scores $\tilde{\boldsymbol{H}}$ to estimate the outer relations (loadings/weights). There are two modes of constructing the measurement model:
- Mode A (reflective), where the arrows point outward from the LV to the OVs, as in our case, see Figure 1. The outer scores are called loadings (λ_{kj}) and estimated by OLS simple linear regression between each observed indicator of the measurement block and the corresponding LV score $\tilde{h}_j$.
- Mode B (formative), where the arrows point inward from the observed variables to the corresponding LVs. The outer estimated scores called weights (w_{kj}) and are calculated by OLS multiple linear regression in which each LV score $\tilde{h}_j$ is regressed on all the observed indicators of the corresponding block.

All the outer estimates λ_{kj}s and w_{kj}s are collected in the updated weight matrix $\hat{\boldsymbol{W}}$.

iv. Using the obtained weight matrix $\hat{\boldsymbol{W}}$ to update the LV scores

$$\boldsymbol{H} = \boldsymbol{N}\hat{\boldsymbol{W}},$$

where $\boldsymbol{H}$ contains the LV scores of the last iteration process.

This stage iterates sequentially from Step i. to Step iv. until convergence. At the convergence, the final LV case values $\boldsymbol{H}$ are obtained as well as the estimated outer matrices $\boldsymbol{C}$ and $\boldsymbol{G}$

$$\boldsymbol{C} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .6759 & 0 \\ .7477 & 0 \\ .7739 & 0 \\ .7093 & 0 \\ 0 & .5326 \\ 0 & .8217 \\ 0 & .7356 \end{pmatrix} \end{array}, \quad \boldsymbol{G} = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .6881 & 0 \\ .7408 & 0 \\ .7574 & 0 \\ .7287 & 0 \\ 0 & .4929 \\ 0 & .8132 \\ 0 & .7206 \end{pmatrix} \end{array}.$$

The matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ contain the loadings that link the latent vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with $\mathbf{X}$ and $\mathbf{Y}$, respectively.

2. The second stage runs the block Cholesky decomposition of Kiiveri et al. two times on the obtained LV case values. The first decomposition is applied on the inverse of the product moment of the covariance matrix of the LV final scores $\boldsymbol{\Sigma}_{\boldsymbol{H}}^{-1}$ that is obtained at the convergence of the Wold algorithm, see Equation (2.1). The resulting block matrix $\boldsymbol{L}$ gives the estimated path coefficient matrices of the inner relations,

$$\boldsymbol{B} = \begin{array}{c} \\ \eta^1 \\ \eta^2 \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} 1 & .107 \\ 0 & 1 \end{pmatrix} \end{array}, \quad \boldsymbol{A} = \begin{array}{c} \\ \eta^1 \\ \eta^2 \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .337 & .156 \\ -.068 & .577 \end{pmatrix} \end{array};$$

whereas, the block matrix $\boldsymbol{D}$ yields the error covariance matrices

$$\boldsymbol{Q} = cov(\boldsymbol{\zeta}) = \begin{pmatrix} .869 & 0 \\ 0 & .663 \end{pmatrix}, \quad \boldsymbol{F} = cov(\boldsymbol{\xi}) = \begin{pmatrix} 1 & -.024 \\ -.024 & 1 \end{pmatrix}.$$

The second decomposition is performed on the inverse of the product moments of the shifted LV score pairs $\boldsymbol{\Sigma}^{-1}_{\boldsymbol{H}_{s,s+1}}$, see Equation (3.2). The resulting block matrix $\boldsymbol{L}^*$ contains

$$\boldsymbol{U} = \begin{matrix} \\ \xi^1 \\ \xi^2 \end{matrix} \begin{matrix} \begin{matrix} \xi^1 & \xi^2 \end{matrix} \\ \begin{pmatrix} 1 & .014 \\ 0 & 1 \end{pmatrix} \end{matrix}, \quad \boldsymbol{V} = \begin{matrix} \\ \xi^1 \\ \xi^2 \end{matrix} \begin{matrix} \begin{matrix} \eta^1 & \eta^2 \end{matrix} \\ \begin{pmatrix} .037 & -.074 \\ .051 & .145 \end{pmatrix} \end{matrix};$$

while, the matrix $\boldsymbol{D}^*$ gives the error covariance matrices

$$\boldsymbol{R} = cov(\boldsymbol{\gamma}) = \begin{pmatrix} .992 & 0 \\ 0 & .984 \end{pmatrix}, \quad \boldsymbol{F}^* = cov(\boldsymbol{\eta}) = \begin{pmatrix} 1 & -.041 \\ -.041 & 1 \end{pmatrix}.$$

At this point, the specified parameter matrices are obtained from the training sample. The estimated matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ show the loadings of each OV on the corresponding LV. The matrix $\boldsymbol{B}$ displays the extent to which girls' views on mobility affect her views on participating in making decisions; while $\boldsymbol{A}$ shows how parental views on girls' mobility and decision making influence their daughters' opinions on these issues. There is a direct effect of parental views on their daughters' opinions in the same domain, i.e., parents who reported a conservative view on girls' participation in making decisions tend to lead their daughters to think alike. The same scenario is true for mobility, where daughters tend to reproduce their parents' views. On the contrary, the effect is small when we consider parental opinions of one domain on their daughters' views of the other domain.

As new cases come one by one at a time sequence ($t = 1, 2, \dots\ T$), instead of re-running the estimation algorithm, we give a recursion to predict the LV case values for the new observation. To do so, the estimated parameter matrices based on the training sample and the Kálmán filtering technique will be used. Theorem 3.1 discusses the recursion from which the optimal prediction of the latent case values is obtained along with the covariance matrix of the prediction error. Specifically, the prediction of $\hat{\boldsymbol{\eta}}_t$ utilizes the estimated $\hat{\boldsymbol{\xi}}_t$ and the new observation $\mathbf{X}_t$, while the new $\mathbf{Y}_t$ and the obtained $\hat{\boldsymbol{\eta}}_t$ are necessary to find $\hat{\boldsymbol{\xi}}_{t+1}$. To start the recursion, the first propagated matrix $\boldsymbol{P}_1$ ought to be initialized from the training sample. Then, the Kálmán gain matrices $\boldsymbol{K}_t$ and $\boldsymbol{M}_t$ are obtained. The succession of calculations follows the order:

$$\boldsymbol{P}_t \to \boldsymbol{K}_t \to \hat{\boldsymbol{\eta}}_t \to \check{\boldsymbol{\eta}}_t \to \boldsymbol{P}^*_t \to \boldsymbol{M}_t \to \check{\boldsymbol{\xi}}_{t+1} \to \hat{\boldsymbol{\xi}}_{t+1} \to \hat{\mathbf{X}}_{t+1} \to \boldsymbol{P}_{t+1}.$$

For $t = 1$, we show the results of the highlighted matrices of the recursion as they are derived in Section 3.1 and Section 3.2.
First stage: $\mathbf{X}_1 \to \hat{\boldsymbol{\eta}}_1$

$$\boldsymbol{P}_1 = \begin{matrix} \begin{matrix} \xi^1 & \xi^2 \end{matrix} \\ \begin{pmatrix} .007 & .005 \\ .005 & .003 \end{pmatrix} \end{matrix},$$

$$\boldsymbol{K}_1 = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \end{array} \\ \begin{pmatrix} -.018 & -.036 & -.036 & -.022 & .022 & .064 & .051 \\ -.013 & -.025 & -.025 & -.015 & .015 & .045 & .036 \end{pmatrix} \end{array},$$

$$\hat{\boldsymbol{\eta}}_1 = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .601 & -.377 \end{pmatrix} \end{array}, \quad \boldsymbol{P}_1^* = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{cc} \eta^1 & \eta^2 \end{array} \\ \begin{pmatrix} .878 & -.070 \\ -.070 & .664 \end{pmatrix} \end{array}.$$

where $\hat{\boldsymbol{\eta}}_1$ show the predicted case values based on the information $\mathbf{X}_1$ of the new observation; and $\boldsymbol{P}_1^*$ is the covariance matrix of the prediction error of $\hat{\boldsymbol{\eta}}_1$.
Second stage: $\mathbf{Y}_1 \to \hat{\boldsymbol{\xi}}_2$

$$\boldsymbol{M}_1 = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 \end{array} \\ \begin{pmatrix} .005 & .016 & .019 & .009 & -.022 & -.042 & -.034 \\ .032 & .011 & .007 & .018 & .041 & .086 & .064 \end{pmatrix} \end{array},$$

$$\hat{\boldsymbol{\xi}}_2 = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} -.595 & -1.460 \end{pmatrix} \end{array}, \quad \boldsymbol{P}_2 = \begin{array}{c} \\ \end{array} \begin{array}{c} \begin{array}{cc} \xi^1 & \xi^2 \end{array} \\ \begin{pmatrix} .993 & -.014 \\ -.014 & .985 \end{pmatrix} \end{array}.$$

where $\hat{\boldsymbol{\xi}}_2$ presents the estimated case values for the exogenous LVs at $t = 2$ based on the new information $\mathbf{Y}_1$. Then the covariance matrix of the prediction error $\hat{\boldsymbol{\xi}}_2$ is obtained. From this we can calculate the propagation matrix $\boldsymbol{P}_2$ to start the recursion once again at $t = 2$ for the next new observation.

The root mean square error ($RMSE$) statistic measures the prediction errors. For the test sample of size 70 observations, we compared the predicted LV case values that are obtained from the Wold algorithm and the filtering technique, simultaneously. Table 1 shows the values of $RSME$. It indicates that the prediction capability of the filtering technique and that of the Wold algorithm are quite homogeneous. Moreover, the difference (in Frobenius norm) between the error covariance and gain matrices in the tth and $(t+1)$th consecutive steps of the recursion are displayed in Table 2. This shows that these matrices are stabilized after the first few steps.

TABLE 1. RMSE for the predictions of the test sample.

Test Obs. at	Wold Prediction (W)				Filtering Prediction (F)			
Sequence "t"	$\hat{\xi}_t^1$	$\hat{\xi}_t^2$	$\hat{\eta}_t^1$	$\hat{\eta}_t^2$	$\hat{\xi}_t^1$	$\hat{\xi}_t^2$	$\hat{\eta}_t^1$	$\hat{\eta}_t^2$
1	-.44005	-.28108	-.25305	-.46325	-.33149	-.52816	-.32865	-.14788
2	.68377	.28108	.84937	-.46325	.89402	.30331	.44740	-.34648
3	.68377	.28108	.84937	-.46325	.35488	.18442	.46531	-.36156
4	.59157	1.50271	.18313	1.18603	.59467	2.43277	.73569	2.05939
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
70	-.53202	-.28108	-.69213	-.46325	-.47413	-.54895	-.59782	-.31266
$RMSE$:	$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(W(LV_t) - F(LV_t))^2}$				.3075	.4033	.3117	.3229

TABLE 2. Consecutive norm for gain and propagation of predictions.

Sequence	Frobenius Norm $\|.\|_F$			
"t"	$\boldsymbol{P}_{t+1} - \boldsymbol{P}_t$	$\boldsymbol{K}_{t+1} - \boldsymbol{K}_t$	$\boldsymbol{P}^*_{t+1} - \boldsymbol{P}^*_t$	$\boldsymbol{M}_{t+1} - \boldsymbol{M}_t$
1	-	-	-	-
2	1.391595	.5148828	.008475387	8.145129e-5
3	7.044589e-7	7.491063e-9	1.094965e-10	1.138605e-12
4	9.341999e-15	1.487013e-16	4.388542e-17	1.357636e-17
5	1.734723e-18	0	0	0
6	0	0	0	0
⋮	⋮	⋮	⋮	⋮
70	0	0	0	0

In sum, the numerical results show a good performance of our proposed algorithm. Once the specified matrices are obtained from the training sample, the Kálmán filtering technique yields an optimal prediction for the LV case values along with the error covariance matrices for the test sample.

5. Discussion and Conclusion

It should be emphasized that the PLS method of Wold is applicable to a given sample, where estimates for the endogeneus variables are given through the exogenous latent ones, and the case values of the LVs are also estimated. The algorithm uses a lot of OLS regressions and so, the estimation of the coefficient matrices is time demanding akin to the block Cholesky decomposition we use. This is the case when we have a long time series with small time intervals or data when the observations come frequently in subsequent order. Our point is that for the new observations, we need not to repeat the whole estimation procedure to obtain the model parameters, but instead we can update the latent variable scores with the help of the new observable data, the estimated matrices, and the Kálmán filtering technique.

In this way, an artificial intelligence is developed. The parameter matrices are estimated from a training sample at the beginning, and the latent variable scores are estimated as observable variables arrive one by one from the test sample. Moreover, there is no need for any distribution assumptions and the data are not necessarily independent. It should be noted that in the possession of a stationary time series, the matrix sequences $\boldsymbol{K}_t$ and $\boldsymbol{M}_t$ (as $t \to \infty$) tend to fixed points of an iteration finding the solution of a matrix Riccati equation (see [4]), but this is the topic of a further research.

Acknowledgement. The research reported in this paper which was carried out at the Budapest University of Technology and Economics has been supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology; also supported by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program). The research was suported by the following project too:

EFOP-3.6.2-16-2017-00015, HU-MATHS-IN for deepening the activity of the Hungarian Industrial and Innovation Network.

References

[1] Haavelmo, T., *The statistical implications of a system of simultaneous equations*, Econometrica, **11**(1943), 1-12.

[2] Jöreskog, K.G., *Structural equation models in the social sciences specification, estimation and testing*, In: "Applications of Statistics", (P.R. Krishnaiah, Ed), North-Holland Publishing Co., 1977, 265-287.

[3] Kálmán, R.E., *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic. Eng., **82D**(1960), 35-45.

[4] Kálmán, R.E., Bucy, R.S., *New results in linear filtering and prediction theory*, Trans. Amer. Soc. Mech. Eng., J. Basic Eng., **83**(1961), 95-108.

[5] Kiiveri, H., Speed, T.P., Carlin, J.B., *Recursive casual models*, J. Aust. Math. Soc., **36**(1984), 30-52.

[6] Rao, C.R., *Linear Statistical Inference and its Applications*, Wiley (1973).

[7] Tenenhaus, M., Esposito Vinzi, V., Chatelinc, Y-L., Lauro, C., *PLS path modeling*, Comput. Statist. Data Anal., **48.1**(2005), 159-205.

[8] Wold, H., *Partial least squares*, In: "Encyclopedia of Statistical Sciences", (Kotz, S., Johnson, N. L., Ed), Wiley, New York, **6**(1985), 581-591.

Marianna Bolla
Budapest University of Technology and Economics,
Institute of Mathematics,
Műegyetem rkp. 3. Budapest 1111, Hungary
e-mail: marib@math.bme.hu

Fatma Abdelkhalek
Budapest University of Technology and Economics,
Institute of Mathematics,
Műegyetem rkp. 3. Budapest 1111, Hungary
Also at Faculty of Commerce,
Assiut University, Egypt
e-mail: fatma@math.bme.hu