DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 15/2019

A flexible state-space model with lagged states and lagged dependent variables: Simulation smoothing

Philipp Hauber
(Kiel Institute for the World Economy)

Christian Schumacher
(Deutsche Bundesbank)

Jiachun Zhang
(Deutsche Bundesbank)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–581–1 (Printversion)
ISBN 978–3–95729–582–8 (Internetversion)

Non-technical summary

Research Question

In many central banks, models with unobserved components are routinely applied for policy analysis and forecasting. When the models are estimated in a Bayesian framework with iterative algorithms, recurrent sampling of the unobserved components given the available data and parameters is necessary. In the literature, this procedure is called simulation smoothing. The computational burden of this sampling step can be considerable, even if modern personal computers are applied.

Contribution

The paper proposes a fast simulation smoother in linear state-space models when observations are missing in the data. The simulation smoother is based on a particular reformulation of the basic state-space form: It includes lagged states in the observation equation and its intercepts depend on lagged observations. In this particular state-space model, the state vector can be kept small, which can make simulation smoothing more efficient computationally. Expanding on the existing literature, we derive the Kalman smoother moments and the simulation smoother algorithm. To illustrate the method, the approach is applied to a large dynamic factor model with many variables. The paper compares the computing time of the simulation smoother sampler to other approaches based on the Kalman filter and smoother.

Results

The proposed simulation smoother is competitive in terms of computational speed. It works faster than the simulation smoothers based on the Kalman filter and smoother without lagged states in the observation equation and without intercepts depending on lagged dependent variables.

Nichttechnische Zusammenfassung

Fragestellung

In vielen Zentralbanken werden Modelle mit unbeobachtbaren Komponenten, sogenannte Zustandsraummodelle, regelmäßig für die Analyse wirtschaftspolitischer Maßnahmen und Prognosen verwendet. Bei der bayesianischen Schätzung solcher Modelle kommen häufig iterative Algorithmen zum Einsatz, in denen wiederholt zufällig aus der stochastischen Verteilung der unbeobachtbaren Komponenten bedingt auf verfügbare Daten und Parameter gezogen wird. Der rechnerische Aufwand dieser Ziehungen kann trotz des Einsatzes moderner Computer beträchtlich sein.

Beitrag

Dieses Papier schlägt ein Verfahren zum Ziehen von unbeobachtbaren Komponenten in linearen Zustandsraummodellen vor, wenn der verwendete Datensatz fehlende Beobachtungen aufweist. Die Methode basiert auf einer bestimmten Umformulierung des Zustandsraummodells: Es enthält verzögerte Zustandsvariablen in der Beobachtungsgleichung, und die Absolutterme hängen von verzögerten Beobachtungen ab. In diesem Zustandsraummodell kann die Dimension des Zustandsvektors klein gehalten werden, was Einsparungen beim Zeitaufwand für die Schätzung erwarten lässt. Die Methode wird anhand eines Faktormodells mit einem großen Datensatz veranschaulicht. Das Papier vergleicht die Rechenzeit des vorgeschlagenen Ansatzes mit anderen Verfahren, die auf dem Kalmanfilter und -glätter beruhen.

Ergebnisse

In Bezug auf die Rechenzeit erweist sich der vorgeschlagene Ansatz als vorteilhaft. Das Verfahren benötigt in verschiedenen Simulationsexperimenten weniger Berechnungszeit als Verfahren, die auf dem Kalmanfilter und -glätter beruhen.

Deutsche Bundesbank Discussion Paper No 15/2019

A flexible state-space model with lagged states and lagged dependent variables: Simulation smoothing*

Philipp Hauber
Kiel Institute for the World Economy

Christian Schumacher
Deutsche Bundesbank

Jiachun Zhang
Deutsche Bundesbank

Abstract

We provide a simulation smoother to a flexible state-space model with lagged states and lagged dependent variables. Qian (2014) has introduced this state-space model and proposes a fast Kalman filter with time-varying state dimension in the presence of missing observations in the data. In this paper, we derive the corresponding Kalman smoother moments and propose an efficient simulation smoother, which relies on mean corrections for unconditional vectors. When applied to a factor model, the proposed simulation smoother for the states is efficient compared to other state-space models without lagged states and/or lagged dependent variables in terms of computing time.

Keywords: State-space model, missing observations, Kalman filter and smoother, simulation smoothing, factor model

JEL classification: C11, C32, C38, C63, C55.

*Contact address: Christian Schumacher, Deutsche Bundesbank, Economic Research Centre, Wilhelm-Epstein-Str. 14, 60431 Frankfurt, Germany. E-Mail: christian.schumacher@bundesbank.de.
Helpful comments were received by Malte Kurz, Elmar Mertens, and seminar participants at the Deutsche Bundesbank. We also thank Hang Qian for kindly providing Kalman filter codes. The computer codes for the simulation smoother were implemented in Matlab 2016a.

1 Introduction

This paper proposes a simulation smoother for a state-space model with lagged states and lagged dependent variables in both the observation and state equations.

Let α_t be an $(p_t \times 1)$ state vector with time-varying dimension p_t, and Y_t be an $(q_t \times 1)$ observation vector, whose dimension q_t can also vary over time. Considering time periods $t = 1, \ldots, n$, define $Y_1^t := \{Y_1, \ldots, Y_t\}$ as the information set of all observations from the first period up to time t. The flexible state-space model by Qian (2014) has the following observation and state equations

$$\alpha_t = f_t(Y_1^{t-1}) + F_t\alpha_{t-1} + \epsilon_t, \tag{1}$$
$$Y_t = g_t(Y_1^{t-1}) + H_t\alpha_t + J_t\alpha_{t-1} + u_t, \tag{2}$$

where the intercepts $f_t(\cdot)$ and $g_t(\cdot)$ are deterministic functions, and ϵ_t, u_t are distributed as

$$\begin{pmatrix} \epsilon_t \\ u_t \end{pmatrix} \sim \mathcal{N}\left[0, \begin{pmatrix} Q_t & S_t \\ S_t' & R_t \end{pmatrix}\right]. \tag{3}$$

The parameters F_t, H_t, J_t, Q_t, R_t, and S_t are potentially time-varying conformable to time variation in the dimensions of α_t and Y_t. The initial state vector α_0 is distributed as $\alpha_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$.

The model differs from standard formulations of state-space models like $\alpha_t = c_t + F_t\alpha_{t-1} + \epsilon_t, Y_t = d_t + H_t\alpha_t + u_t$ with respect to the explicit dependence on lagged dependent variables by $f_t(Y_1^{t-1})$ and $g_t(Y_1^{t-1})$, as well as the lagged state $J_t\alpha_{t-1}$. This specification is efficient computationally for Kalman filtering, as it helps to keep the dimension of the state-vector small when some observations are missing. Qian (2014) discusses these advantages in three kinds of models: an autoregressive moving-average (ARMA) model, a factor model, and autoregressive conditional heteroscedasticity (ARCH) models.

In this paper, we derive the Kalman smoother moments for the model (1)-(3) and propose an efficient simulation smoother, which relies on mean corrections for unconditional vectors as in Durbin and Koopman (2002). Qian (2014) provides the Kalman filter recursions, but not the smoother or simulation smoother recursions. The smoother for the model without lagged dependent variables is discussed in depth by Nimark (2015) and Kurz (2018).

The paper proceeds as follows: The Kalman filter including its initialization is provided in Section 2. The Kalman smoother and the simulation smoother are discussed in Section 3. In Section 4, we discuss how the simulation smoother can be applied to a factor model. We also compare two alternative simulation smoothers, both relying on different state-space representations: A representation with time-invariant state dimension and neither lagged dependent variables nor lagged states in the observation equation, and a time-varying dimension state-space factor model as proposed by Jungbacker, Koopman, and van der Wel (2011), which considers lagged dependent variables, but not lagged states in the observation equation. Section 5 provides a simulation experiment to compare the computational speed of the three simulation smoothers. Section 6 concludes. In what follows, I is an identity matrix whose dimension may vary and a 0 in matrices indicates a matrix of zeros.

2 Kalman filter and initialization

Given the assumptions for the initial conditions and the noise terms, Qian (2014) derives the Kalman filter recursions for the state-space model (1)-(3). The results are summarized in Algorithm 1, including the moments of the state update $\alpha_t | Y_1^t \sim \mathcal{N}(\hat{\alpha}_{t|t}, P_{t|t})$ for $t = 1, \ldots, n$.

Algorithm 1 (Kalman filter by Qian (2014))**.** Consider the unconditional initial value of the state, $\alpha_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$. Y_1^0 is assumed to be empty, hence, the mean and covariance of the state in period $t = 0$ conditional on initial information are equal to $\hat{\alpha}_{0|0} = \mu_0, P_{0|0} = \Sigma_0$, respectively. For periods $t = 1, \ldots, n$, iterate through the following steps:

Step 1: For prediction, compute the moments of the joint distribution of the states and observed variables conditional on information from period $t-1$, which is defined as

$$\begin{pmatrix} \alpha_t \\ Y_t \end{pmatrix} | Y_1^{t-1} \sim \mathcal{N}\left[\begin{pmatrix} \hat{\alpha}_{t|t-1} \\ \hat{Y}_{t|t-1} \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & L_{t|t-1} \\ L'_{t|t-1} & D_{t|t-1} \end{pmatrix}\right],$$

with

$$\hat{\alpha}_{t|t-1} = f_t(Y_1^{t-1}) + F_t \hat{\alpha}_{t-1|t-1}, \tag{4}$$

$$\hat{Y}_{t|t-1} = g_t(Y_1^{t-1}) + H_t \hat{\alpha}_{t|t-1} + J_t \hat{\alpha}_{t-1|t-1}, \tag{5}$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F'_t + Q_t, \tag{6}$$

$$\begin{aligned} D_{t|t-1} = {} & H_t P_{t|t-1} H'_t + R_t + J_t P_{t-1|t-1} J'_t \\ & + H_t F_t P_{t-1|t-1} J'_t + J_t P_{t-1|t-1} F'_t H'_t + H_t S_t + S'_t H'_t, \end{aligned} \tag{7}$$

$$L_{t|t-1} = P_{t|t-1} H'_t + F_t P_{t-1|t-1} J'_t + S_t. \tag{8}$$

Step 2: Conditional on the additional observation Y_t in period t, the update of the state is distributed as $\alpha_t | Y_1^t \sim \mathcal{N}(\hat{\alpha}_{t|t}, P_{t|t})$ with moments

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + L_{t|t-1} D_{t|t-1}^{-1} (Y_t - \hat{Y}_{t|t-1}), \tag{9}$$

$$P_{t|t} = P_{t|t-1} - L_{t|t-1} D_{t|t-1}^{-1} L'_{t|t-1}. \tag{10}$$

The recursions in Algorithm 1 are based on the assumption that the initial condition is fixed and known. Alternatively, one can explicitly consider uncertainty regarding the initial observation vector Y_0 and state. For this purpose, assume that the intercept functions $f(\cdot)$ and $g(\cdot)$ are linear functions of the previous period's observed variables only, in particular, $f_t(Y_1^{t-1}) = f_t Y_{t-1}$ and $g_t(Y_1^{t-1}) = g_t Y_{t-1}$. In this case, one can specify the initial condition

$$\begin{pmatrix} \alpha_0 \\ Y_0 \end{pmatrix} \sim \mathcal{N}(\mu_0^*, \Sigma_0^*) \tag{11}$$

and reformulate the model in $t = 1$ by temporarily augmenting the state vector with the initial observations

$$\alpha_1^* = F_1^* \alpha_0^* + \epsilon_1^*, \quad Y_1 = H_1^* \alpha_1^*, \tag{12}$$

where

$$\alpha_1^* = \begin{pmatrix} \alpha_1 \\ Y_1 \end{pmatrix}, \quad \alpha_0^* = \begin{pmatrix} \alpha_0 \\ Y_0 \end{pmatrix}, \tag{13}$$

$$F_1^* = \begin{pmatrix} F_1 & f_1 \\ H_1F_1 + J_1 & g_1 + H_1 f_1 \end{pmatrix}, \quad \epsilon_1^* = \begin{pmatrix} \epsilon_1 \\ H_1\epsilon_1 + u_1 \end{pmatrix}, \tag{14}$$

$$Q_1^* = \begin{pmatrix} Q_1 & Q_1H_1' + S_1 \\ H_1Q_1 + S_1' & H_1Q_1H_1' + R_1 + S_1'H_1' + H_1S_1 \end{pmatrix}, \tag{15}$$

$$H_1^* = \begin{pmatrix} 0 & I \end{pmatrix}. \tag{16}$$

These equations constitute a state space model without measurement errors, and the prediction and update steps are the conventional Kalman filter recursions, initialized with $\hat{\alpha}_{0|0}^* = \mu_0^*, P_{0|0} = \Sigma_0^*$. Specifically, they are given by

$$\hat{\alpha}_{1|0}^* = F_1^*\hat{\alpha}_{0|0}^*, \tag{17}$$

$$\hat{Y}_{1|0} = H_1^*\hat{\alpha}_{1|0}^*, \tag{18}$$

$$P_{1|0} = F_1^*P_{0|0}F_1^{*\prime} + Q_1^*, \tag{19}$$

$$D_{1|0} = H_1^*P_{1|0}H_1^{*\prime}, \tag{20}$$

$$L_{1|0} = P_{1|0}H_1^{*\prime}, \tag{21}$$

and

$$\hat{\alpha}_{1|1}^* = \hat{\alpha}_{1|0}^* + L_{1|0}D_{1|0}^{-1}(Y_1 - \hat{Y}_{1|0}), \tag{22}$$

$$P_{1|1} = P_{1|0} - L_{1|0}D_{1|0}^{-1}L_{1|0}', \tag{23}$$

respectively.

In period $t = 2$, switch from the standard to the flexible state space model by removing the observations in the augmented state vector. Thus, the transition equation is given by

$$\alpha_2 = F_2^*\alpha_1^* + \epsilon_2 \tag{24}$$

with $F_2^* = \begin{pmatrix} F_2 & f_2 \end{pmatrix}$ and the measurement equation

$$Y_2 = g_2Y_1 + H_2\alpha_2 + J_2^*\alpha_1^* + u_2 \tag{25}$$

with $J_2^* = \begin{pmatrix} J_2 & 0 \end{pmatrix}$. The prediction step is then

$$\hat{\alpha}_{2|1} = F_2^*\hat{\alpha}_{1|1}^*, \tag{26}$$

$$\hat{Y}_{2|1} = g_2Y_1 + H_2\hat{\alpha}_{2|1} + J_2^*\hat{\alpha}_{1|1}^*, \tag{27}$$

$$P_{2|1} = F_2^*P_{1|1}F_2^{*\prime} + Q_1, \tag{28}$$

$$\begin{aligned} D_{2|1} = {} & H_2P_{2|1}H_2{}' + R_2 + J_2^*P_{1|1}J_2^{*\prime} \\ & + H_2F_2^*P_{1|1}J_2^{*\prime} + J_2^*P_{1|1}F_2^{*\prime}H_2{}' + H_2S_2 + S_2'H_2', \end{aligned} \tag{29}$$

$$L_{2|1} = P_{2|1}H_2{}' + F_2^*P_{1|1}J_2^{*\prime} + S_2. \tag{30}$$

Updating $\hat{\alpha}_{2|1}$ and $P_{2|1}$ proceeds as in Algorithm 1. The same holds for periods $t \geq 3$.

Higher-order lags of the dependent variables in the intercepts can be considered in a similar way.

3 Kalman smoother and simulation smoothing

Our goal is to derive a simulation smoother for the state-space model (1)-(3). Extending the Kalman filter results by Qian (2014), we provide the smoothed state distribution $\alpha_t|Y_1^n \sim \mathcal{N}(\hat{\alpha}_{t|n}, P_{t|n})$. The mean and the covariance of the smoothed state vector, $\hat{\alpha}_{t|n} = \mathbb{E}(\alpha_t|Y_1^n)$ and $P_{t|n} = \mathrm{Var}(\alpha_t|Y_1^n)$, are provided in Proposition 1. Proofs can be found in Appendix A.1.

Proposition 1 (Moments of smoothed state vector based on $\hat{\alpha}_{t|t-1}$ and $P_{t|t-1}$)**.** *Let*

$$\begin{aligned}
A_t &:= [F_t - L_{t|t-1}D_{t|t-1}^{-1}(H_tF_t + J_t)]', \\
B_t &:= Q_t[I - L_{t|t-1}D_{t|t-1}^{-1}H_t]' - S_t(L_{t|t-1}D_{t|t-1}^{-1})', \\
C_t &:= (H_tF_t + J_t)', \\
v_t &:= Y_t - \hat{Y}_{t|t-1},
\end{aligned}$$

Moreover, define $r_{n+1} = 0$, and

$$r_t = C_tD_{t|t-1}^{-1}v_t + A_tr_{t+1}$$

for $t = 1, ..., n$. The mean of the smoothed state vector $\hat{\alpha}_{t|n}$ is given by

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t-1} + (Q_tH_t' + S_t)\, D_{t|t-1}^{-1}v_t + F_tP_{t-1|t-1}r_t + B_tr_{t+1}. \tag{31}$$

Let $N_{n+1} = 0$ and

$$N_t = C_tD_{t|t-1}^{-1}C_t' + A_tN_{t+1}A_t'$$

for $t = 1, ..., n$. The covariance matrix of the smoothed state $P_{t|n}$ is given by

$$\begin{aligned}
P_{t|n} = P_{t|t-1} &- (Q_tH_t' + S_t)\, D_{t|t-1}^{-1}\, (Q_tH_t' + S_t)' - (Q_tH_t' + S_t)\, D_{t|t-1}^{-1}(F_tP_{t-1|t-1}C_t)' \\
&- F_tP_{t-1|t-1}C_tD_{t|t-1}^{-1}\, (Q_tH_t' + S_t)' - F_tP_{t-1|t-1}N_tP_{t-1|t-1}'F_t' - B_tN_{t+1}B_t'.
\end{aligned} \tag{32}$$

The smoother moments in Proposition 1 are based on the predicted state moments $\hat{\alpha}_{t|t-1}$ and $P_{t|t-1}$ given information in period $t-1$ for $t = 1, ..., n$. In Proposition 2, we also derive the smoother moments based on the updated state moments $\hat{\alpha}_{t|t}$ and $P_{t|t}$ given information in period t. Compared to Proposition 1, the expressions are simpler and thus preferable in practice for computational reasons. Note that the moments in Proposition 2 are the same as in Kurz (2018), Section 4.2, apart from the time-varying system matrices in our model. Interestingly, the model in Kurz (2018) does contain lagged states in the observation equation, but not lagged dependent variables in the state and observation equations as in the model considered in this paper. Proposition 2 shows that the existence of lagged dependent variables does not change the formulas for the smoothed moments. The reason is that the lagged dependent variables change only the formulas for the prediction step, namely, Step 1 in Algorithm 1, by additional information from period

$t-1$. The subsequent computation of quantities for the update steps and smoothing are not affected. Proofs can be found in Appendix A.2.

Proposition 2 (Moments of smoothed state vector based on $\hat{\alpha}_{t|t}$ and $P_{t|t}$)**.** *Define A_t, C_t, and v_t as in Proposition 1. Moreover, define $r_n^* = 0$, and*

$$r_t^* = C_{t+1} D_{t+1|t}^{-1} v_{t+1} + A_{t+1} r_{t+1}^* \tag{33}$$

for $t = 1, ..., n-1$. The mean of the smoothed state vector $\hat{\alpha}_{t|n}$ is given by

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t} + P_{t|t} r_t^*. \tag{34}$$

Let $N_n^* = 0$ and

$$N_t^* = C_{t+1} D_{t+1|t}^{-1} C_{t+1}' + A_{t+1} N_{t+1}^* A_{t+1}'$$

for $t = 1, ..., n-1$. The covariance matrix of the smoothed state $P_{t|n}$ is given by

$$P_{t|n} = P_{t|t} - P_{t|t} N_t^* P_{t|t}. \tag{35}$$

Concerning the relationship between Proposition 1 and Proposition 2, note that $r_t^* = r_{t+1}$ and $N_t^* = N_{t+1}$ for $t = 1, ..., n$. It can be shown that the means (31) and (34) as well as the covariances in (32) and (35) are equal, see Appendix A.3 for details.

Given the recursions for the mean of the smoothed state $\mathbb{E}(\alpha_t | Y_1^n)$, the simulation smoother is provided in Algorithm 2, which adopts Algorithm 2 of Durbin and Koopman (2002). A formal justification for Algorithm 2 is provided in Appendix A.4.

Algorithm 2 (Simulation smoother)**.** Define $\alpha = (\alpha_1', \ldots, \alpha_n')'$ and $Y = (Y_1', \ldots, Y_n')'$. Follow three steps to obtain a draw for α conditional on Y:

Step 1: Simulate α^+ and Y^+ by means of recursion (1) and (2), where the recursion is initialized by a draw from $\alpha_0^+ \sim N(\mu_0, \Sigma_0)$ or augmented by Y_0 as in (11).

Step 2: Compute $\hat{\alpha} = \mathbb{E}(\alpha|Y)$ and $\hat{\alpha}^+ = \mathbb{E}(\alpha^+|Y^+)$ by means of the Kalman filter and smoother using Proposition 2 based on observed and simulated data, respectively.

Step 3: Take $\tilde{\alpha} = \hat{\alpha} - \hat{\alpha}^+ + \alpha^+$. $\tilde{\alpha}$ is a draw from the distribution of α conditional on Y.

Note that the intercepts $f_t(\cdot)$ and $g_t(\cdot)$ are not exogenous, but rather endogenous functions of Y^+ in Step 1 of Algorithm 2. In Step 2, the intercepts differ for the two smoother runs depending on Y^+ and Y. This has two implications. First, as the intercepts in the model differ conditional on Y^+ and on Y, we have to compute $\hat{\alpha}^+$ and $\hat{\alpha}$ separately and perform the mean adjustment separately. It is not possible to directly compute the mean differential by running the Kalman smoother only once with $Y^* = Y - Y^+$. Second, this also implies that there is no need to reset the initial conditions regardless of μ_0^* (μ_0) and Σ_0^* (Σ_0) as in Jarociński (2015).

4 Application: Factor model with missing observations

To illustrate the simulation smoother in Algorithm 2, we choose an application to a factor model with vector autoregressive (VAR) dynamics for the factors and autoregressive (AR) idiosyncratic components. We focus on this kind of model, as it has been discussed extensively in the literature (Jungbacker et al., 2011; Banbura and Modugno, 2014; Qian, 2014), in particular, when observations in the data are missing.

The factor model explains the $(N \times 1)-$dimensional vector of variables $x_t := (x_{1,t}, x_{2,t}, \dots, x_{N,t})'$ in time period t according to

$$x_t = \lambda \eta_t + \epsilon_t, \tag{36}$$

$$\eta_t = \phi \eta_{t-1} + u_{\eta,t}, \quad \epsilon_t = \psi \epsilon_{t-1} + u_{\epsilon,t}. \tag{37}$$

The $(r \times 1)-$dimensional vector of factors is denoted as η_t, and λ is the $(N \times r)-$dimensional matrix of factor loadings. The factor representation (36) holds for $t = 1, \dots, n$. The factors follow a VAR(1) process with the $(r \times r)-$dimensional lag parameter matrix ϕ. The factor VAR residuals are distributed as $u_{\eta,t} \sim \mathcal{N}(0_{r\times 1}, \omega_\eta)$. The idiosyncratic components collected in the $(N \times 1)-$dimensional vector $\epsilon_t := (\epsilon_{1,t}, \epsilon_{2,t}, \dots, \epsilon_{N,t})'$ each follow AR(1) processes such that the $(N \times N)-$dimensional coefficient matrix ψ is diagonal containing the AR(1) lag parameters ψ_i for $i = 1, \dots, N$ on the main diagonal. The idiosyncratic disturbances are distributed as $u_{\epsilon,t} \sim \mathcal{N}(0_{N\times 1}, \omega_\epsilon)$, where ω_ϵ is also diagonal with diagonal elements $\omega_{\epsilon,i}$ for $i = 1, \dots, N$. We assume that $u_{\eta,t}$ and $u_{\epsilon,t}$ are mutually independent and that the VAR and AR processes in (37) are stationary. In addition, the equations in (37) are defined for time periods $t = 2, \dots, n$. For $t = 1$, let $\eta_1 \sim \mathcal{N}(\eta_{1|0}, \vartheta_{\eta,1|0})$ and $\epsilon_1 \sim \mathcal{N}(\epsilon_{1|0}, \vartheta_{\epsilon,1|0})$, respectively. The means of the distributions are set equal to their unconditional mean, which is zero in our case, so $\eta_{1|0} = 0_{r\times 1}$ and $\epsilon_{1|0} = 0_{N\times 1}$, respectively. As the equations in (37) are stationary, $\vartheta_{\eta,1|0}$ and $\vartheta_{\epsilon,1|0}$ are set equal to the implied unconditional covariances. We collect all the model parameters in $\theta = \{\lambda, \phi, \psi, \omega_\eta, \omega_\epsilon, \eta_{1|0}, \vartheta_{\eta,1|0}, \epsilon_{1|0}, \vartheta_{\epsilon,1|0}\}$.

Given a complete data set, a parsimonious state-space representation of the factor model can be obtained by quasi-differencing the observations, for example, by $\tilde{x}_t = x_t - \psi x_{t-1}$ to remove the idiosyncratic autocorrelation. The state vector only contains η_t in this case. However, as discussed in detail by Jungbacker et al. (2011), this representation is not valid in the presence of missing observations. A valid alternative is a time-invariant state-space representation with a state vector containing both the factors and the idiosyncratic components as used in Banbura and Modugno (2014). More efficient representations can be obtained by allowing for time-varying dimensions of the state vector. We discuss these alternative models in turn, starting with the state-space model from (1)-(3) including lagged states and lagged dependent variables.

4.1 State-space model with lagged states and lagged dependent variables

In order to apply simulation smoothing as described in Algorithm 2 to the factor model (36) and (37), a particular time-varying state-space form is chosen. To obtain a short state vector, the observation equation is quasi-differenced first. We obtain the modified observation equation $x_t = \psi x_{t-1} + \lambda\eta_t - \psi\lambda\eta_{t-1} + u_{\epsilon,t}$, which has lagged dependent variables and lagged states on the right-hand side. When stacking contemporaneous and lagged states into $(\eta_t', \eta_{t-1}')'$, this observation equation together with the state equation $\eta_t = \phi\eta_{t-1} + u_{\eta,t}$ yields a state-space model with state dimension $2r$. Additionally, Qian (2014) removes the contemporary factor from the observation equation by using the factor VAR $\eta_t = \phi\eta_{t-1} + u_{\eta,t}$. After this transformation, the observation equation contains only lagged states according to

$$\begin{aligned} x_t(o_t) &= \psi(o_t, o_{t-1}) x_{t-1}(o_{t-1}) \\ &\quad + \begin{pmatrix} G(o_t, :) & \psi(o_t, m_{t-1}) \end{pmatrix} \begin{pmatrix} \eta_{t-1} \\ x_{t-1}(m_{t-1}) \end{pmatrix} + v_t(o_t), \end{aligned} \tag{38}$$

where $G = \lambda\phi - \psi\lambda$ and $v_t = \lambda u_{\eta,t} + u_{\epsilon,t}$. The transition equation is written as

$$\begin{aligned} \begin{pmatrix} \eta_t \\ x_t(m_t) \end{pmatrix} &= \begin{pmatrix} 0 \\ \psi(m_t, o_{t-1}) x_{t-1}(o_{t-1}) \end{pmatrix} \\ &\quad + \begin{pmatrix} \phi & 0 \\ G(m_t, :) & \psi(m_t, m_{t-1}) \end{pmatrix} \begin{pmatrix} \eta_{t-1} \\ x_{t-1}(m_{t-1}) \end{pmatrix} + \begin{pmatrix} u_{\eta,t} \\ v_t(m_t) \end{pmatrix}. \end{aligned} \tag{39}$$

In the equations above, o_t and m_t are logical indices indicating the observed and missing entries in x_t at time t, respectively. For any vector v, $v(o_t)$ and $v(m_t)$ select the elements in v, which correspond to the observed and missing entries in x_t, respectively. For a matrix G, $G(:, o_t)$ denotes column selection, $G(o_t, :)$ denotes row selection, in both examples with respect to observed values in period t. By using two index matrices such as $\psi(m_t, o_{t-1})$, we can select columns and rows, in this example, the rows corresponding to missing values in period t and the columns corresponding to observed values in $t-1$.

When applying the simulation smoother in Algorithm 2, we have to consider the intercepts in the following way. At time t, we have the information of $x_{t-1}(o_{t-1})$, which can be used to correct both intercept terms. In this case, no extra information needs to be saved for constructing the state-space model. Dividing both intercept terms with $x_{t-1}(o_{t-1})$, then multiplying them with new draw $x^+_{t-1}(o_{t-1})$ leads to the right constant terms.

Below, we denote the approach by 'Lagged states and lagged dependent variables'.

4.2 State-space model with lagged dependent variables only

Jungbacker et al. (2011) discuss a state-space model with time-varying state-dimension and lagged states in both the observation and state equation, but no lagged states in the observation equation.

Denote by K_{t-1} a selection matrix of ones and zeros such that for a vector v

$$\begin{pmatrix} v_t(o_t m_{t-1}) \\ v_t(m_t m_{t-1}) \end{pmatrix} = K_{t-1} v_t(m_{t-1}),$$

where the double index for row selection indicates a logical AND in the following. For example, indexing by $o_t m_{t-1}$ picks those elements in v_t, which are observed in period t, but not observed in period $t-1$. The state-space model in Jungbacker et al. (2011) can be expressed as

$$\begin{aligned}
\begin{pmatrix} x_t(o_t o_{t-1}) \\ x_t(o_t m_{t-1}) \end{pmatrix} &= \begin{pmatrix} \psi(o_t o_{t-1}, o_t o_{t-1}) x_{t-1}(o_t o_{t-1}) \\ 0 \end{pmatrix} \\
&\quad + \begin{pmatrix} \lambda(o_t o_{t-1}, :) & -\psi(o_t o_{t-1}, o_t o_{t-1})\lambda(o_t o_{t-1}, :) & 0 & 0 \\ \lambda(o_t m_{t-1}, :) & 0 & I & 0 \end{pmatrix} \begin{pmatrix} \eta_t \\ \eta_{t-1} \\ \epsilon_t(o_t m_{t-1}) \\ \epsilon_t(m_t) \end{pmatrix} \\
&\quad + \begin{pmatrix} u_{\epsilon,t}(o_t o_{t-1}) \\ 0 \end{pmatrix}, \qquad (40)
\end{aligned}$$

$$\begin{aligned}
\begin{pmatrix} \eta_t \\ \eta_{t-1} \\ \epsilon_t(o_t m_{t-1}) \\ \epsilon_t(m_t) \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ \psi(m_t o_{t-1}, m_t o_{t-1}) x_{t-1}(m_t o_{t-1}) \end{pmatrix} \\
&+ \begin{pmatrix} \phi & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{t-1}\psi(m_{t-1}, m_{t-1}) \\ -\psi(m_t o_{t-1}, m_t o_{t-1})\lambda(m_t o_{t-1}, :) & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \eta_{t-1} \\ \eta_{t-2} \\ \epsilon_{t-1}(o_{t-1} m_{t-2}) \\ \epsilon_{t-1}(m_{t-1}) \end{pmatrix} \\
&+ \begin{pmatrix} u_{\eta,t} \\ 0 \\ K_{t-1} u_{\epsilon,t}(m_{t-1}) \\ u_{\epsilon,t}(m_t o_{t-1}) \end{pmatrix}. \qquad (41)
\end{aligned}$$

Compared to the state-space model (38) and (39), there are no lagged states in the observation equation (40). The state vector always includes the factors and the first-order lags of the factors. Idiosyncratic terms are included in a time-varying fashion to tackle missing values in the current and previous period.

The factor model representation (40) and (41) is a state-space form with time-varying dimensions of the variables and system matrices. Hence, we can apply the simulation smoother by Durbin and Koopman (2002) in a similar way as in Algorithm 2. Special care is needed when considering the intercepts in Step 2 of Algorithm 2. Note that at time t, we have the information of

$$\begin{pmatrix} x_{t-1}(o_{t-1} o_{t-2}) \\ x_{t-1}(o_{t-1} m_{t-2}) \end{pmatrix} = x_{t-1}(o_{t-1}),$$

and need to correct the intercept terms with the help of

$$\begin{pmatrix} x_{t-1}(o_t o_{t-1}) \\ x_{t-1}(m_t o_{t-1}) \end{pmatrix} = x_{t-1}(o_{t-1}).$$

We have to reorder $x_{t-1}(o_{t-1})$ with the missing value information at time $t-2$ and time t. This can be implemented when defining the time-varying system matrices of the state-space model. Dividing both constant terms with $x_{t-1}(o_t o_{t-1})$ and $x_{t-1}(m_t o_{t-1})$ respectively, then multiplying them with new draw $x^+_{t-1}(o_t o_{t-1})$ and $x^+_{t-1}(m_t o_{t-1})$ leads to the correct constant terms.

Below, we denote the approach by 'Lagged dependent variables only'.

4.3 State-space model with time-invariant state dimension

The factor model in (36) and (37) can be cast in another state-space form by including the idiosyncratic components in the state vector, yielding

$$x_t = \begin{pmatrix} \lambda & I \end{pmatrix} \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix}, \tag{42}$$

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 0 & \psi \end{pmatrix} \begin{pmatrix} \eta_{t-1} \\ \epsilon_{t-1} \end{pmatrix} + \begin{pmatrix} u_{\eta,t} \\ u_{\epsilon,t} \end{pmatrix}, \tag{43}$$

which has been used in the literature by Banbura and Modugno (2014), for example. To consider missing observations for simulation smoothing, we follow Durbin and Koopman (2012), Section 4.10, and assume that we can observe x_t in period t only partly and define the selection matrix W_t such that $x_t(o_t) = W_t x_t$. By pre-multiplying the observation equation (42) for each t, we can apply the simulation smoother in Algorithm 2. When simulating Y^+ in Step 1, the case of time-invariant state dimension necessitates to set the values in Y^+ to be missing at the same positions as there are missing values in Y.

In this model, the dimension of the state vector is constant over time, namely $r + N$. In the other two specifications, the dimension of the state depends on the presence of missing data, and can be considerable smaller the lesser observations are missing.

In the simulations below, we denote the approach by 'Time-invariant state dimension'.

5 Simulation exercise

We consider data-generating processes (DGP) that differ with respect to the sample size T with a set of two values $\{100, 200\}$, the number of variables N with the alternatives $\{50, 100, 200\}$, and the set $\{4, 16\}$ for the number of factors r. Concerning the available data for estimation, we assume that 50% of observations are missing, $\kappa = 0.5$. The missing observations are distributed randomly across the indexes (i, t). Overall we have 12 different experiments of different 4−tuples (T, N, r, κ).

The DGP is defined by the factor model in (36) and (37). The elements of the loading matrix λ are each sampled independently from a Normal distribution according to $\lambda_{ij} \sim \mathcal{N}(0, 1/r^2)$. The AR coefficient of the idiosyncratic components follows a Normal distribution $\psi_i \sim \mathcal{N}(0.5, 0.01)$, the variance of the disturbance in the AR(1) equation is equal to $(1 - \psi_i^2)$. The VAR coefficient matrix implies a recursive dynamic structure. It is upper diagonal with $\phi_{i,i} = \gamma/1^2$ for $i = 1, \dots, r$ on the main diagonal, $\phi_{i,i+1} = \gamma/2^2$ for $i = 1, \dots, r-1$ on the first superdiagonal, up to $\phi_{1,r} = \gamma/r^2$ for the top-right element in ϕ, and γ is distributed as $\gamma \sim \mathcal{N}(0.8, 0.01)$. The remaining elements in ϕ are equal to zero. The off-diagonal elements of the covariance matrix ω_η are set to zero. The diagonal elements of ω_η are set such that the variances of the factors in population are each equal to one. Given this set of parameters θ, we sample η_{DGP}, ϵ_{DGP}, and x_{DGP} from the DGP (36) and (37). We randomly set κNT observations to missing values, yielding the data set x^o used for sampling states and missing observations. Given the data set $x = x^o$, we draw $K = 200$ times new parameters $\theta^{(k)}$, and $\eta^{(k)} \sim p(\eta|x = x^o, \theta^{(k)})$ using the different samplers discussed in the previous section.

In Figure 1, we compare the overall computing time elapsed to draw the $K = 200$ samples from $p(\eta|x = x^o, \theta)$ in seconds. In Panel A of Figure 1, the results for $r = 4$ factors are shown, in Panel B for $r = 16$ factors. In Panel A, the results show for all simulation designs with respect to (T, N) a clear pattern. The model with lagged states and lagged dependent variables by Qian (2014) performs better than the sampler with time-invariant state dimension, and the model based on lagged dependent variables only. The model based on lagged dependent variables only always outperforms the model with time-invariant state dimension. When the number of factors is increased to $r = 16$, see Panel B, the results are very similar.

6 Conclusions

In this paper, we provide a simulation smoother for a flexible state-space model with lagged states and lagged dependent variables. Qian (2014) has derived the Kalman filter recursions for this model. We derive the corresponding Kalman smoother moments and propose a simulation smoother based on Durbin and Koopman (2002). When applied to a factor model, the proposed simulation smoother is highly efficient in terms of computing time compared to other approaches where 1) neither lagged states are considered in the observation equation in the model nor lagged dependent variables or 2) only lagged dependent variables are considered.

A Appendix

A.1 Proof of Proposition 1

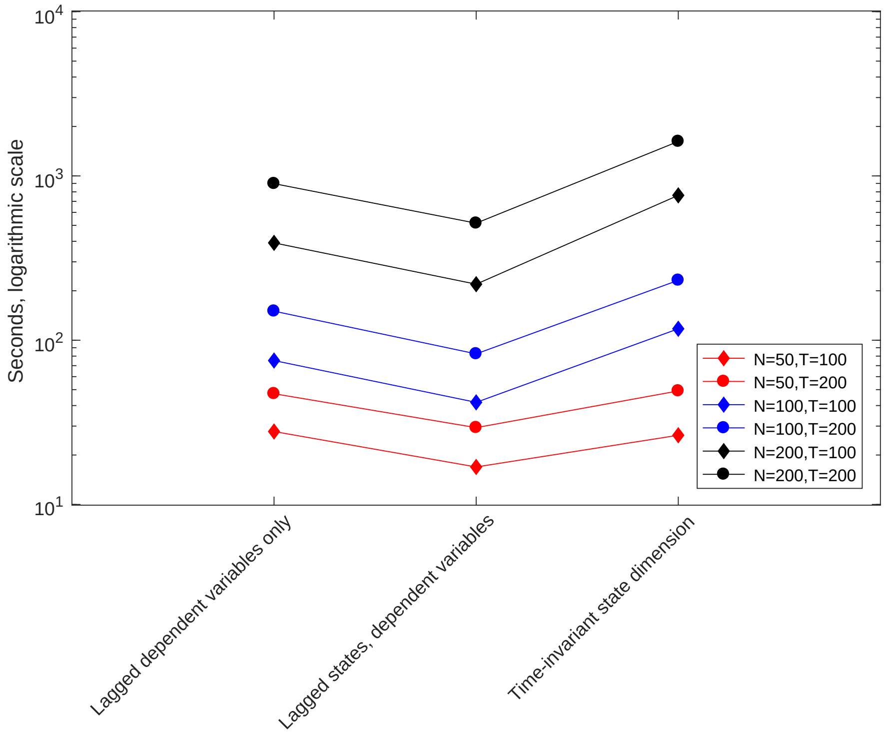
A.1.1 Mean of the smoothed state vector

To derive the mean of the smoothed state vector $\hat{\alpha}_{t|n}$, we adopt the method in Durbin and Koopman (2012), equation (4.31), according to

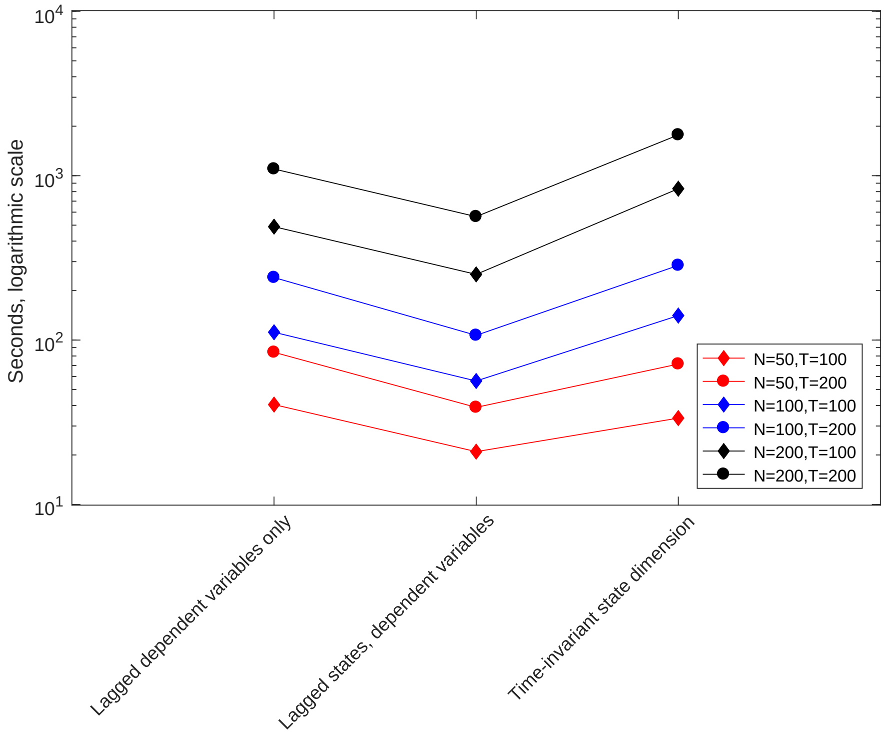
$$\begin{aligned}\hat{\alpha}_{t|n} &= \mathbb{E}(\alpha_t | Y_1^n) = \mathbb{E}(\alpha_t | Y_1^{t-1}, v_{t:n}) \\ &= \hat{\alpha}_{t|t-1} + \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j | Y_1^{t-1}) \mathrm{Var}(v_j | Y_1^{t-1})^{-1} v_j, \end{aligned} \tag{44}$$

Figure 1: Computing time

A. Number of factors $r = 4$



B. Number of factors $r = 16$



where $v_t := Y_t - \hat{Y}_{t|t-1}$. Denote $x_t := \alpha_t - \hat{\alpha}_{t|t-1}$ and $x_{t-1|t-1} := \alpha_{t-1} - \hat{\alpha}_{t-1|t-1}$. Moreover, we define

$$A_t := [F_t - L_{t|t-1}D^{-1}_{t|t-1}(H_tF_t + J_t)]', \tag{45}$$
$$B_t := Q_t[I - L_{t|t-1}D^{-1}_{t|t-1}H_t]' - S_t(L_{t|t-1}D^{-1}_{t|t-1})', \tag{46}$$
$$C_t := (H_tF_t + J_t)'. \tag{47}$$

First note that

$$x_t = \alpha_t - \hat{\alpha}_{t|t-1} \overset{(4),(1)}{=} F_t x_{t-1|t-1} + \epsilon_t \tag{48}$$

and

$$\begin{aligned} v_t &= Y_t - \hat{Y}_{t|t-1} \\ &\overset{(2),(5)}{=} H_t(\alpha_t - \hat{\alpha}_{t|t-1}) + J_t(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}) + u_t \\ &= H_t x_t + J_t x_{t-1|t-1} + u_t \\ &\overset{(48)}{=} (H_tF_t + J_t)x_{t-1|t-1} + H_t\epsilon_t + u_t. \end{aligned} \tag{49}$$

Given the state prediction from the Kalman filter recursion (9), as well as (48) and (49) above, we can derive $x_{t|t}$ and $x_{t+1|t+1}$ according to

$$\begin{aligned} x_{t|t} &= \alpha_t - \hat{\alpha}_{t|t} \\ &\overset{(9)}{=} \alpha_t - \hat{\alpha}_{t|t-1} - L_{t|t-1}D^{-1}_{t|t-1}v_t \\ &= x_t - L_{t|t-1}D^{-1}_{t|t-1}v_t \\ &\overset{(49),(48)}{=} F_t x_{t-1|t-1} + \epsilon_t - L_{t|t-1}D^{-1}_{t|t-1}(H_tF_t + J_t)x_{t-1|t-1} \\ &\qquad - L_{t|t-1}D^{-1}_{t|t-1}H_t\epsilon_t - L_{t|t-1}D^{-1}_{t|t-1}u_t \\ &= [F_t - L_{t|t-1}D^{-1}_{t|t-1}(H_tF_t + J_t)]x_{t-1|t-1} + [I - L_{t|t-1}D^{-1}_{t|t-1}H_t]\epsilon_t \\ &\qquad - L_{t|t-1}D^{-1}_{t|t-1}u_t \\ &= A_t' x_{t-1|t-1} + [I - L_{t|t-1}D^{-1}_{t|t-1}H_t]\epsilon_t - L_{t|t-1}D^{-1}_{t|t-1}u_t, \end{aligned} \tag{50}$$

and, analogously,

$$x_{t+1|t+1} = A'_{t+1}x_{t|t} + (I - L_{t+1|t}D^{-1}_{t+1|t}H_{t+1})\epsilon_{t+1} - L_{t+1|t}D^{-1}_{t+1|t}u_{t+1}. \tag{51}$$

For $j = t, ..., n$, given (49) and since v_j is centered, we can determine the elements in the decomposition (44)

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_j|Y_1^{t-1}) &= \mathbb{E}\left[\alpha_t v_j'|Y_1^{t-1}\right] - \mathbb{E}\left[\alpha_t|Y_1^{t-1}\right] \cdot 0 \\ &= \mathbb{E}\left[\alpha_t x'_{j-1|j-1}(H_jF_j + J_j)' + \alpha_t\epsilon_j' H_j' + \alpha_t u_j'|Y_1^{t-1}\right] \\ &= \mathbb{E}\left[\alpha_t x'_{j-1|j-1}|Y_1^{t-1}\right](H_jF_j + J_j)' + \mathbb{E}\left[\alpha_t\epsilon_j'|Y_1^{t-1}\right]H_j' + \mathbb{E}\left[\alpha_t u_j'|Y_1^{t-1}\right], \end{aligned}$$

where

$$\mathbb{E}\left[\alpha_t \epsilon_j' | Y_1^{t-1}\right] \overset{(1)}{=} \begin{cases} Q_t, & j = t \\ 0, & j = t+1, \ldots, n \end{cases}, \qquad \mathbb{E}\left[\alpha_t u_j' | Y_1^{t-1}\right] \overset{(2)}{=} \begin{cases} S_t, & j = t \\ 0, & j = t+1, \ldots, n \end{cases}.$$

We obtain the expectation terms $\mathbb{E}[\alpha_t x_{j-1|j-1}' | Y_1^{t-1}]$ for $j = t, t+1, ..., n$, respectively,

$$\begin{aligned} \mathbb{E}\left[\alpha_t x_{t-1|t-1}' | Y_1^{t-1}\right] &\overset{(1)}{=} \mathbb{E}\left[(f_t(Y_1^{t-1}) + F_t \alpha_{t-1} + \epsilon_t)(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1})' | Y_1^{t-1}\right] \\ &= F_t P_{t-1|t-1}, \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[\alpha_t x_{t|t}' | Y_1^{t-1}\right] &\overset{(50)}{=} \mathbb{E}\Big[\alpha_t \Big(A_t' x_{t-1|t-1} \\ &\qquad + \left[I - L_{t|t-1} D_{t|t-1}^{-1} H_t\right] \epsilon_t - L_{t|t-1} D_{t|t-1}^{-1} u_t\Big)' | Y_1^{t-1}\Big] \\ &= \mathbb{E}\left[\alpha_t x_{t-1|t-1}' | Y_1^{t-1}\right] A_t \\ &\qquad + \mathbb{E}\left[\alpha_t \epsilon_t' | Y_1^{t-1}\right] \left[I - L_{t|t-1} D_{t|t-1}^{-1} H_t\right]' \\ &\qquad - \mathbb{E}\left[\alpha_t u_t' | Y_1^{t-1}\right] (L_{t|t-1} D_{t|t-1}^{-1})' \\ &= F_t P_{t-1|t-1} A_t \\ &\qquad + Q_t \left[I - L_{t|t-1} D_{t|t-1}^{-1} H_t\right]' - S_t (L_{t|t-1} D_{t|t-1}^{-1})' \\ &= F_t P_{t-1|t-1} A_t + B_t, \end{aligned}$$

$$\begin{aligned} \mathbb{E}\left[\alpha_t x_{t+1|t+1}' | Y_1^{t-1}\right] &\overset{(51)}{=} \mathbb{E}\left[\alpha_t x_{t|t}' | Y_1^{t-1}\right] A_{t+1} + \mathbb{E}\left[\alpha_t \epsilon_{t+1}' | Y_1^{t-1}\right] (I - L_{t+1|t} D_{t+1|t}^{-1} H_{t+1})' \\ &\qquad - \mathbb{E}\left[\alpha_t u_{t+1}' | Y_1^{t-1}\right] (L_{t+1|t} D_{t+1|t}^{-1})' \\ &= F_t P_{t-1|t-1} A_t A_{t+1} + B_t A_{t+1}, \end{aligned}$$

and, finally,

$$\mathbb{E}\left[\alpha_t x_{n-1|n-1}' | Y_1^{t-1}\right] = F_t P_{t-1|t-1} A_t A_{t+1} ... A_{n-1} + B_t A_{t+1} ... A_{n-1}.$$

Notice that

$$\mathrm{Var}(v_t | Y_1^{t-1}) = \mathrm{Var}(Y_t - \hat{Y}_{t|t-1} | Y_1^{t-1}) = D_{t|t-1},$$

we consider the $\hat{\alpha}_{t|n}$ from (44). For $t = n, j = n$, we have

$$\hat{\alpha}_{n|n} = \hat{\alpha}_{n|n-1} + \left(F_n P_{n-1|n-1} C_n + Q_n H_n' + S_n\right) D_{n|n-1}^{-1} v_n.$$

For $t = n-1, j = n-1, n$,

$$\begin{aligned} \hat{\alpha}_{n-1|n} = \hat{\alpha}_{n-1|n-2} &+ \left(F_{n-1} P_{n-2|n-2} C_{n-1} + Q_{n-1} H_{n-1}' + S_{n-1}\right) D_{n-1|n-2}^{-1} v_{n-1} \\ &+ (F_{n-1} P_{n-2|n-2} A_{n-1} + B_{n-1}) C_n D_{n|n-2}^{-1} v_n. \end{aligned}$$

For $t = n-2, j = n-2, n-1, n$,

$$\begin{aligned}\hat{\alpha}_{n-2|n} = \hat{\alpha}_{n-2|n-3} &+ (F_{n-2}P_{n-3|n-3}C_{n-2} + Q_{n-2}H'_{n-2} + S_{n-2})D^{-1}_{n-2|n-3}v_{n-2} \\ &+ \left(F_{n-2}P_{n-3|n-3}A_{n-2} + B_{n-2}\right)C_{n-1}D^{-1}_{n-1|n-3}v_{n-1} \\ &+ (F_{n-2}P_{n-3|n-3}A_{n-2}A_{n-1} + B_{n-2}A_{n-1})C_nD^{-1}_{n|n-3}v_n.\end{aligned}$$

In general,

$$\begin{aligned}\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t-1} &+ (Q_tH'_t + S_t)\,D^{-1}_{t|t-1}v_t + F_tP_{t-1|t-1}C_tD^{-1}_{t|t-1}v_t \\ &+ (F_tP_{t-1|t-1}A_t + B_t)C_{t+1}D^{-1}_{t+1|t-1}v_{t+1} \\ &+ \left(F_tP_{t-1|t-1}A_tA_{t+1} + B_tA_{t+1}\right)C_{t+2}D^{-1}_{t+2|t-1}v_{t+2} \\ &+ ... + \left(F_tP_{t-1|t-1}A_tA_{t+1}...A_{n-1} + B_tA_{t+1}...A_{n-1}\right)C_nD^{-1}_{n|t-1}v_n,\end{aligned}$$

where the definitions (45), (46), and (47) have been used.

Concerning $D_{j|t-1}$ for $j = t, \ldots, n$, we follow Durbin and Koopman (2012) and use

$$D_{j|t-1} = D_{j|j-1}.$$

For $j = t$ this is trivial. For $j = t+1, ..., n$, this assumption holds, since

$$\begin{aligned}D_{j|t-1} &= \mathrm{Var}(v_j|Y_1, v_2, ..., v_{t-1}), \\ D_{j|j-1} &= \mathrm{Var}(v_j|Y_1, v_2, ..., v_{t-1}, v_t, ..., v_{j-1}),\end{aligned}$$

and $v_2, ..., v_t, ..., v_{j-1}$ are mutually independent of each other by definition. Hence additional information of $v_t, ..., v_{j-1}$ does not improve the forecast of v_j.

To simplify the recursions, we first consider the part containing A_t and define $r_n = C_nD^{-1}_{n|n-1}v_n$, then for $t = 1, ..., n-1$,

$$r_t = C_tD^{-1}_{t|t-1}v_t + A_tr_{t+1}. \tag{52}$$

Regarding terms containing B_t, define $l_n = 0$, then for $t = 1, \ldots, n-1$,

$$l_t = C_tD^{-1}_{t|t-1}v_t + A_tl_{t+1}. \tag{53}$$

Note that $l_{t-1} = r_t$. Define $r_{n+1} = l_n = 0$, then

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t-1} + (Q_tH'_t + S_t)\,D^{-1}_{t|t-1}v_t + F_tP_{t-1|t-1}r_t + B_tr_{t+1},$$

which is equal to (31) and completes the the proof of Proposition 1 with respect to the mean of the smoothed state.

A.1.2 Covariance of the smoothed state vector

To derive the covariance of the smoothed state vector $P_{t|n}$, we adopt the method in Durbin and Koopman (2012), Section 4.4.3, according to

$$\begin{aligned} P_{t|n} &= \mathrm{Var}(\alpha_t|Y_1^n) = \mathrm{Var}(\alpha_t|Y_1^{t-1}, v_{t:n}) \\ &= P_{t|t-1} - \sum_{j=t}^{n} \mathrm{Cov}(\alpha_t, v_j|Y_1^{t-1}) \mathrm{Var}(v_j|Y_1^{t-1})^{-1} \mathrm{Cov}(\alpha_t, v_j|Y_1^{t-1})'. \end{aligned} \tag{54}$$

For $t = n, j = n$, we have

$$P_{n|n} = P_{n|n-1} - \left(F_n P_{n-1|n-1} C_n + Q_n H_n' + S_n\right) D_{n|n-1}^{-1} \left(F_n P_{n-1|n-1} C_n + Q_n H_n' + S_n\right)'.$$

For $t = n-1, j = n-1, n$,

$$\begin{aligned} P_{n-1|n} = P_{n-1|n-2} & \\ - \left(F_{n-1} P_{n-2|n-2} C_{n-1} + Q_{n-1} H_{n-1}' + S_{n-1}\right) D_{n-1|n-2}^{-1} & \\ \cdot \left(F_{n-1} P_{n-2|n-2} C_{n-1} + Q_{n-1} H_{n-1}' + S_{n-1}\right)' & \\ - (F_{n-1} P_{n-2|n-2} A_{n-1} + B_{n-1}) C_n D_{n|n-2}^{-1} C_n' (F_{n-1} P_{n-2|n-2} A_{n-1} + B_{n-1})'. & \end{aligned}$$

For $t = n-2, j = n-2, n-1, n$,

$$\begin{aligned} P_{n-2|n} = P_{n-2|n-3} & \\ - (F_{n-2} P_{n-3|n-3} C_{n-2} + Q_{n-2} H_{n-2}' + S_{n-2}) D_{n-2|n-3}^{-1} & \\ \cdot (F_{n-2} P_{n-3|n-3} C_{n-2} + Q_{n-2} H_{n-2}' + S_{n-2})' & \\ - \left(F_{n-2} P_{n-3|n-3} A_{n-2} + B_{n-2}\right) C_{n-1} D_{n-1|n-3}^{-1} C_{n-1}' \left(F_{n-2} P_{n-3|n-3} A_{n-2} + B_{n-2}\right)' & \\ - (F_{n-2} P_{n-3|n-3} A_{n-2} A_{n-1} + B_{n-2} A_{n-1}) C_n D_{n|n-3}^{-1} C_n' & \\ \cdot (F_{n-2} P_{n-3|n-3} A_{n-2} A_{n-1} + B_{n-2} A_{n-1})'. & \end{aligned}$$

In general for $t = 1, \ldots, n$,

$$\begin{aligned} P_{t|n} &= P_{t|t-1} - (Q_t H_t' + S_t) D_{t|t-1}^{-1} (Q_t H_t' + S_t)' - (Q_t H_t' + S_t) D_{t|t-1}^{-1} (F_t P_{t-1|t-1} C_t)' \\ &\quad - F_t P_{t-1|t-1} C_t D_{t|t-1}^{-1} (Q_t H_t' + S_t)' - F_t P_{t-1|t-1} C_t D_{t|t-1}^{-1} C_t' P_{t-1|t-1}' F_t' \\ &\quad - (F_t P_{t-1|t-1} A_t + B_t) C_{t+1} D_{t+1|t-1}^{-1} C_{t+1}' (F_t P_{t-1|t-1} A_t + B_t)' \\ &\quad - \left(F_t P_{t-1|t-1} A_t A_{t+1} + B_t A_{t+1}\right) C_{t+2} D_{t+2|t-1}^{-1} C_{t+2}' \left(F_t P_{t-1|t-1} A_t A_{t+1} + B_t A_{t+1}\right)' \\ &\quad - \ldots - \left(F_t P_{t-1|t-1} A_t A_{t+1} \ldots A_{n-1} + B_t A_{t+1} \ldots A_{n-1}\right) C_n D_{n|t-1}^{-1} C_n' \\ &\qquad \cdot \left(F_t P_{t-1|t-1} A_t A_{t+1} \ldots A_{n-1} + B_t A_{t+1} \ldots A_{n-1}\right)', \end{aligned}$$

where the definitions (45), (46), and (47) have been used.

We again use $D_{j|t-1} = D_{j|j-1}$ for $j = t+1, \ldots, n$ as in Durbin and Koopman (2012). Let $N_{n+1} = 0$, $N_n = C_n D_{n|n-1}^{-1} C_n'$, and for $t = 1, \ldots, n-1$,

$$N_t = C_t D_{t|t-1}^{-1} C_t' + A_t N_{t+1} A_t', \tag{55}$$

then

$$P_{t|n} = P_{t|t-1} - (Q_t H_t' + S_t) D_{t|t-1}^{-1} (Q_t H_t' + S_t)' - (Q_t H_t' + S_t) D_{t|t-1}^{-1} (F_t P_{t-1|t-1} C_t)'$$
$$- F_t P_{t-1|t-1} C_t D_{t|t-1}^{-1} (Q_t H_t' + S_t)' - F_t P_{t-1|t-1} N_t P_{t-1|t-1}' F_t' - B_t N_{t+1} B_t', \quad (56)$$

which is equal to (32) and completes the proof of Proposition 1.

A.2 Proof of Proposition 2

A.2.1 Mean of the smoothed state vector

We start this proof by writing the mean of the smoothed state as a function of the mean of the updated state, as in Kurz (2018),

$$\begin{aligned} \hat{\alpha}_{t|n} &= \mathbb{E}(\alpha_t | Y_1^n) = \mathbb{E}(\alpha_t | Y_1^t, v_{t+1:n}) \\ &= \hat{\alpha}_{t|t} + \sum_{j=t+1}^{n} \mathrm{Cov}(\alpha_t, v_j | Y_1^t) \mathrm{Var}(v_j | Y_1^t)^{-1} v_j. \end{aligned} \quad (57)$$

Note that for $j = t+1, ..., n$, $D_{j|t} = D_{j|j-1}$, hence $\mathrm{Var}(v_j|Y_1^t)^{-1} = D_{j|j-1}^{-1}$. Moreover,

$$\begin{aligned} \mathrm{Cov}(\alpha_t, v_j | Y_1^t) &= \mathbb{E}\left[\alpha_t v_j' | Y_1^t\right] - \mathbb{E}\left[\alpha_t | Y_1^t\right] \cdot 0 \\ &\overset{(49)}{=} \mathbb{E}\left[\alpha_t x_{j-1|j-1}' (H_j F_j + J_j)' + \alpha_t \epsilon_j' H_j' + \alpha_t u_j' | Y_1^t\right] \\ &= \mathbb{E}\left[\alpha_t x_{j-1|j-1}' | Y_1^t\right] (H_j F_j + J_j)' + \mathbb{E}\left[\alpha_t \epsilon_j' | Y_1^t\right] H_j' + \mathbb{E}\left[\alpha_t u_j' | Y_1^t\right], \end{aligned}$$

where $\mathbb{E}\left[\alpha_t \epsilon_j' | Y_1^t\right] = 0$ and $\mathbb{E}\left[\alpha_t u_j' | Y_1^t\right] = 0$, for $j = t+1, ..., n$ by definition. Consider now the term $\mathbb{E}\left[\alpha_t x_{j-1|j-1}' | Y_1^t\right]$. For $j = t+1$, $\mathbb{E}\left[\alpha_t x_{t|t}' | Y_1^t\right] = P_{t|t}$. For $j = t+2$,

$$\begin{aligned} &\mathbb{E}\left[\alpha_t x_{t+1|t+1}' | Y_1^t\right] \\ \overset{(51)}{=}\ &\mathbb{E}\left[\alpha_t x_{t|t}' | Y_1^t\right] A_{t+1} + \mathbb{E}\left[\alpha_t \epsilon_{t+1}' | Y_1^t\right] [I - L_{t+1|t} D_{t+1|t}^{-1} H_{t+1}]' - \mathbb{E}\left[\alpha_t u_{t+1}' | Y_1^t\right] [L_{t+1|t} D_{t+1|t}^{-1}]' \\ =\ &P_{t|t} A_{t+1}. \end{aligned}$$

In general,

$$\mathrm{Cov}(\alpha_t, v_j | Y_1^t) = P_{t|t} A_{t+1} ... A_{j-1} C_j, \quad (58)$$

and (57) becomes

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t} + \sum_{j=t+1}^{n} P_{t|t} A_{t+1} ... A_{j-1} C_j D_{j|j-1}^{-1} v_j. \quad (59)$$

For $t = n-1$, $j = n$,

$$\hat{\alpha}_{n-1|n} = \hat{\alpha}_{n-1|n-1} + P_{n-1|n-1} C_n D_{n|n-1}^{-1} v_n.$$

For $t = n-2$, $j = n, n-1$,

$$\hat{\alpha}_{n-2|n} = \hat{\alpha}_{n-2|n-2} + P_{n-2|n-2}A_{n-1}C_nD^{-1}_{n|n-1}v_n + P_{n-2|n-2}C_{n-1}D^{-1}_{n-1|n-2}v_{n-1}.$$

Let $r^*_t = C_{t+1}D^{-1}_{t+1|t}v_{t+1} + A_{t+1}r^*_{t+1}$, and $r^*_n = 0$, then

$$\hat{\alpha}_{t|n} = \hat{\alpha}_{t|t} + P_{t|t}r^*_t. \tag{60}$$

A.2.2 Covariance of the smoothed state vector

As in Kurz (2018), we can write the covariance of the smoothed state as a function of the covariance of the updated state according to

$$\begin{aligned} P_{t|n} &= \mathrm{Var}(\alpha_t|Y^n_1) = \mathrm{Var}(\alpha_t|Y^t_1, v_{t+1:n}) \\ &= P_{t|t} - \sum_{j=t+1}^{n} \mathrm{Cov}(\alpha_t, v_j|Y^t_1)\mathrm{Var}(v_j|Y^t_1)^{-1}\mathrm{Cov}(\alpha_t, v_j|Y^t_1)'. \end{aligned} \tag{61}$$

With (58) we have

$$P_{t|n} = P_{t|t} - \sum_{j=t+1}^{n} P_{t|t}A_{t+1}...A_{j-1}C_jD^{-1}_{j|j-1}\left(P_{t|t}A_{t+1}...A_{j-1}C_j\right)'.$$

For $t = n-1$, $j = n$,

$$P_{n-1|n} = P_{n-1|n-1} - P_{n-1|n-1}C_nD^{-1}_{n|n-1}C'_nP_{n-1|n-1}.$$

For $t = n-2$, $j = n, n-1$,

$$\begin{aligned} P_{n-2|n} \;=\; & P_{n-2|n-2} - P_{n-2|n-2}A_{n-1}C_nD^{-1}_{n|n-1}C'_nA'_{n-1}P_{n-2|n-2} \\ & - P_{n-2|n-2}C_{n-1}D^{-1}_{n-1|n-2}C'_{n-1}P_{n-2|n-2}. \end{aligned}$$

Let $N^*_t = C_{t+1}D^{-1}_{t+1|t}C'_{t+1} + A_{t+1}N^*_{t+1}A'_{t+1}$, and $N^*_n = 0$, then

$$P_{t|n} = P_{t|t} - P_{t|t}N^*_tP_{t|t}. \tag{62}$$

A.3 Equivalence of the smoother means in Proposition 1 and Proposition 2

As for the equivalence of the smoother means in Proposition 1 and Proposition 2, the case $t = n$ is clear and the coming proof is for $t = n-1, \ldots, 1$. We start with (34), using the definition of C_t in (47),

$$\begin{aligned}
\hat{\alpha}_{t|n} &= \hat{a}_{t|t} + P_{t|t} r_t^* \\
&\overset{(9)}{=} \hat{\alpha}_{t|t-1} + L_{t|t-1} D_{t|t-1}^{-1} v_t + P_{t|t} r_t^* \\
&\overset{(8),(6)}{=} \hat{\alpha}_{t|t-1} + \left[F_t P_{t-1|t-1} F_t' H_t' + Q_t H_t' + F_t P_{t-1|t-1} J_t' + S_t\right] D_{t|t-1}^{-1} v_t + P_{t|t} r_t^* \\
&= \hat{\alpha}_{t|t-1} + [Q_t H_t' + S_t] D_{t|t-1}^{-1} v_t + F_t P_{t-1|t-1} \underbrace{[H_t F_t + J_t]'}_{=C_t} D_{t|t-1}^{-1} v_t + P_{t|t} r_t^*.
\end{aligned}$$

Comparing this expression with (31) by using the update of r_t, we have

$$\begin{aligned}
\hat{\alpha}_{t|n} &= \hat{\alpha}_{t|t-1} + (Q_t H_t' + S_t) D_{t|t-1}^{-1} v_t + F_t P_{t-1|t-1} r_t + B_t r_{t+1} \\
&= \hat{\alpha}_{t|t-1} + (Q_t H_t' + S_t) D_{t|t-1}^{-1} v_t + F_t P_{t-1|t-1} \left[C_t D_{t|t-1}^{-1} v_t + A_t r_{t+1}\right] + B_t r_{t+1} \\
&= \hat{\alpha}_{t|t-1} + (Q_t H_t' + S_t) D_{t|t-1}^{-1} v_t + F_t P_{t-1|t-1} C_t D_{t|t-1}^{-1} v_t + \left[F_t P_{t-1|t-1} A_t + B_t\right] r_{t+1}.
\end{aligned}$$

Note that $r_t^* = r_{t+1}$. Thus, we only have to show that

$$P_{t|t} \overset{!}{=} F_t P_{t-1|t-1} A_t + B_t. \tag{63}$$

By (6) and (10), the left-hand side of (63) is equal to

$$P_{t|t} = F_t P_{t-1|t-1} F_t' + Q_t - L_{t|t-1} D_{t|t-1}^{-1} L_{t|t-1}'.$$

By definition of A_t and B_t, the right-hand side of (63) is equal to

$$\begin{aligned}
& F_t P_{t-1|t-1} A_t + B_t \\
=\;& F_t P_{t-1|t-1} \left[F_t - L_{t|t-1} D_{t|t-1}^{-1} (H_t F_t + J_t)\right]' + Q_t [I - L_{t|t-1} D_{t|t-1}^{-1} H_t]' \\
& \quad - S_t (L_{t|t-1} D_{t|t-1}^{-1})' \\
=\;& F_t P_{t-1|t-1} F_t' + Q_t - \left[F_t P_{t-1|t-1} (H_t F_t + J_t)' + Q_t H_t' + S_t\right] (L_{t|t-1} D_{t|t-1}^{-1})' \\
\overset{(8),(6)}{=}\;& F_t P_{t-1|t-1} F_t' + Q_t - L_{t|t-1} D_{t|t-1}^{-1} L_{t|t-1}'.
\end{aligned}$$

Thus, the left-hand side and the right-hand side of (63) are identical, which proves the equivalence of the means (31) and (34). The proof for the covariances (32) and (35) works analogously.

A.4 Verifying the mean and the covariance of the state sampled using the simulation smoother in Algorithm 2

Algorithm 2 adopts the simulation smoother by Durbin and Koopman (2002) to the state-space model with lagged states in the observation equation and lagged dependent variables in the state and observation equation. Compared to Durbin and Koopman (2002), we provide a more detailed justification of the simulation smoother. In particular, we check the validity of the simulation smoother in Algorithm 2 by analyzing the moments of the simulated state $\tilde{\alpha}$ given the observed data $Y_1, ..., Y_n$ as realizations of the corresponding random variables. We assume that $f_t(Y_1^{t-1})$ and $g_t(Y_1^{t-1})$ are linear functions of $Y_1, ..., Y_{t-1}$, and stack the states and observations according to $\alpha = (\alpha_1', \ldots, \alpha_n')'$ and $Y = (Y_1', \ldots, Y_n')'$. As the state-space model is linear and the disturbances are normally distributed, the joint vector of states and observables is also multivariate normal with

$$\begin{pmatrix} \alpha \\ Y \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \mu_\alpha \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha Y} \\ \Sigma_{Y\alpha} & \Sigma_{YY} \end{pmatrix}\right], \tag{64}$$

where means μ_α, μ_Y, variances and covariances $\Sigma_{\alpha\alpha}$, $\Sigma_{\alpha Y}$, $\Sigma_{Y\alpha}$, Σ_{YY} are functions of the system matrices $f_t(\cdot)$, $g_t(\cdot)$, F_t, H_t, J_t, Q_t, R_t, S_t for $t = 1, \ldots, n$, and the initial conditions of the state-space model.

Given the joint normal distribution, we can derive the mean and the covariance of the state α conditional on Y according to

$$\mathbb{E}[\alpha|Y] = \mu_\alpha + \Sigma_{\alpha Y}\Sigma_{YY}^{-1}(Y - \mu_Y), \tag{65}$$

$$\mathrm{Var}[\alpha|Y] = \Sigma_{\alpha\alpha} - \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{Y\alpha}. \tag{66}$$

To check the validity of the simulation smoother, we have to show that $\mathbb{E}[\tilde{\alpha}|Y] = \mathbb{E}[\alpha|Y]$ and $\mathrm{Var}[\tilde{\alpha}|Y] = \mathrm{Var}[\alpha|Y]$. Note that from the definition of Algorithm 2, the stacked vector of α^+ and simulated data Y^+ has the independent identical distribution as the stacked vector of α and Y.

Concerning the mean, we have

$$\begin{aligned}\mathbb{E}[\tilde{\alpha}|Y] &= \mathbb{E}[\mathbb{E}(\alpha|Y) - \mathbb{E}(\alpha^+|Y^+) + \alpha^+|Y] \\ &= \mathbb{E}[\mathbb{E}(\alpha|Y)|Y] - \mathbb{E}[\mathbb{E}(\alpha^+|Y^+)|Y] + \mathbb{E}[\alpha^+|Y] \\ &= \mathbb{E}(\alpha|Y) - \mathbb{E}(\alpha^+) + \mathbb{E}(\alpha^+) = \mathbb{E}(\alpha|Y),\end{aligned}$$

where $\mathbb{E}[\mathbb{E}(\alpha|Y)|Y] = \mathbb{E}(\alpha|Y)$ due to the Tower property, $\mathbb{E}[\mathbb{E}(\alpha^+|Y^+)|Y] = \mathbb{E}(\alpha^+)$ and $\mathbb{E}[\alpha^+|Y] = \mathbb{E}(\alpha^+)$ since both Y^+ and α^+ are independent of Y.

Concerning the covariance, note that due to the Tower property,

$$\mathrm{Var}[\mathbb{E}(\alpha|Y)|Y] = \mathbb{E}[(\mathbb{E}(\alpha|Y) - \mathbb{E}[\mathbb{E}(\alpha|Y)|Y])^2|Y] = \mathbb{E}[(\mathbb{E}(\alpha|Y) - \mathbb{E}(\alpha|Y))^2|Y] = 0.$$

Let $Z := -\mathbb{E}(\alpha^+|Y^+) + \alpha^+$, which is independent of Y, then

$$\begin{aligned}\mathrm{Var}[\mathbb{E}(\alpha^+|Y^+) - \alpha^+|Y] &= \mathrm{Var}[Z|Y] \\ &= \mathbb{E}[(Z - \mathbb{E}[Z|Y])^2|Y] = \mathbb{E}[(Z - \mathbb{E}[Z])^2|Y] = \mathbb{E}[(Z - \mathbb{E}[Z])^2] \\ &= \mathrm{Var}[Z] = \mathrm{Var}[\mathbb{E}(\alpha^+|Y^+) - \alpha^+].\end{aligned}$$

Moreover,

$$\begin{aligned}\mathrm{Var}(\mathbb{E}(\alpha^+|Y^+)) &= \mathrm{Var}(\mu_\alpha + \Sigma_{\alpha Y}\Sigma_{YY}^{-1}(Y^+ - \mu_Y)) \\ &= \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{YY}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}' = \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}',\end{aligned}$$

$$\mathrm{Cov}(\mathbb{E}(\alpha^+|Y^+), \alpha^+) = \mathrm{Cov}(\mu_\alpha + \Sigma_{\alpha Y}\Sigma_{YY}^{-1}(Y^+ - \mu_Y), \alpha^+) = \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}'.$$

Together we obtain

$$\begin{aligned}
\mathrm{Var}[\tilde{\alpha}|Y] &= \mathrm{Var}[\mathbb{E}(\alpha|Y) - \mathbb{E}(\alpha^+|Y^+) + \alpha^+|Y] \\
&= \mathrm{Var}[\mathbb{E}(\alpha|Y)|Y] + \mathrm{Var}[\mathbb{E}(\alpha^+|Y^+) - \alpha^+|Y] \\
&= 0 + \mathrm{Var}(\mathbb{E}(\alpha^+|Y^+)) + \mathrm{Var}(\alpha^+) - 2\mathrm{Cov}(\mathbb{E}(\alpha^+|Y^+), \alpha^+) \\
&= \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}' + \Sigma_{\alpha\alpha} - 2\Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}' \\
&= \Sigma_{\alpha\alpha} - \Sigma_{\alpha Y}\Sigma_{YY}^{-1}\Sigma_{\alpha Y}'.
\end{aligned}$$

As the state moments from the simulation smoother and the moments from the theoretical distribution coincide, we can conclude that the simulation smoother provides an unbiased sample with correct covariance.

References

Banbura, M. and M. Modugno (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics 29*(1), 133–160.

Durbin, J. and S. Koopman (2012). *Time Series Analysis by State Space Methods, Second Edition.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–616.

Jarociǹski, M. (2015). A note on implementing the durbin and koopman simulation smoother. *Computational Statistics and Data Analysis 91*, 1–3.

Jungbacker, B., S. J. Koopman, and M. van der Wel (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control 35*(8), 1358–1368.

Kurz, M. S. (2018). A note on low-dimensional kalman smoothers for systems with lagged states in the measurement equation. *Economics Letters 168*, 42–45.

Nimark, K. P. (2015). A low dimensional kalman filter for systems with lagged states in the measurement equation. *Economics Letters 127*, 10–13.

Qian, H. (2014). A flexible state space model and its applications. *Journal of Time Series Analysis 35*(2), 79–88.