Institut für Volkswirtschaftslehre

Universität Augsburg

Volkswirtschaftliche Diskussionsreihe

An Augmented Steady-State Kalman Filter to Evaluate the Likelihood of Linear and Time-Invariant State-Space Models

Johannes Huber

Beitrag Nr. 343, April 2022

An Augmented Steady-State Kalman Filter to Evaluate the Likelihood of Linear and Time-Invariant State-Space Models

Johannes Huber[a]

[a]University of Augsburg, Department of Economics, Universitätsstraße 16, 86159 Augsburg, Germany, johannes.huber@wiwi.uni-augsburg.de

April 7, 2022

JEL classification:
C18, C63, E20
Keywords:
Kalman filter, DSGE, Bayesian estimation, Maximum-likelihood estimation, Computational techniques

Abstract

We propose a modified version of the augmented Kalman filter (AKF) to evaluate the likelihood of linear and time-invariant state-space models (SSMs). Unlike the regular AKF, this augmented steady-state Kalman filter (ASKF), as we call it, is based on a steady-state Kalman filter (SKF). We show that to apply the ASKF, it is sufficient that the SSM at hand is stationary. We find that the ASKF can significantly reduce the computational burden to evaluate the likelihood of medium- to large-scale SSMs, making it particularly useful to estimate dynamic stochastic general equilibrium (DSGE) models and dynamic factor models. Tests using a medium-scale DSGE model, namely the 2007 version of the Smets and Wouters model, show that the ASKF is up to five times faster than the regular Kalman filter (KF). Other competing algorithms, such as the Chandrasekhar recursion (CR) or a univariate treatment of multivariate observation vectors (UKF), are also outperformed by the ASKF in terms of computational efficiency.

Acknowledgment: I am grateful to my Ph.D. advisor Prof. Alfred Maußner for initial ideas and ensuing discussions. I also thank Prof. Burkhard Heer, Dr. Daniel Fehrle, Dr. Christopher Heiberger and Vasilij Konysev for valuable comments.

1 Introduction

Since their introduction in the 1980s, dynamic stochastic general equilibrium (DSGE) models have become a cornerstone of modern macroeconomics. While the first class of DSGE models mainly consisted of small-scale models with a handful of equations and only a few shocks (e.g., Hansen, 1985; King et al., 1988), the complexity of these models has increased significantly over the past decades (see, e.g., Leeper et al., 2010; Gadatsch et al., 2016; Drygalla et al., 2020). In particular, New-Keynesian models, such as those of Christiano et al. (2005) or Smets and Wouters (2003, 2007), are no longer used only for academic purposes but also for monetary policy analysis.[1] A popular approach to specify the parameters of a (log-) linearized DSGE model is to treat its policy function as a linear (and time-invariant) state-space model (SSM), and estimate this SSM using likelihood-based methods (e.g., Ireland, 2004; An and Schorfheide, 2007; Chari et al., 2007). However, as the complexity of the model increases, the repeated evaluation of the likelihood function can become time-consuming.[2] This paper proposes an algorithm to evaluate the likelihood of linear and time-invariant SSMs. We find that this augmented steady-state Kalman filter (ASKF), as we call it, can significantly reduce the time required to evaluate the likelihood of (log-) linearized DSGE models, such as the one introduced by Smets and Wouters (2007). Although we focus mainly on DSGE models in this paper, the ASKF may also be useful for estimating other linear and time-invariant SSMs such as, e.g., dynamic factor models.

There are two likelihood-based approaches to estimate the parameters of DSGE models, namely the frequentist and the Bayesian approach. The frequentist approach considers the set of unknown parameters as fixed and estimates them by maximum-likelihood. The number of likelihood evaluations within this approach remains manageable for a limited amount of unknown parameters and a well-shaped likelihood function. However, even for well-identified models, due to the curse of dimensionality, maximization of the likelihood function often becomes a difficult task as the dimension of the parameter space rises. Thus, problems with a high-dimensional parameter space often require global search routines, such as simulated annealing, to locate the global maxima (see, e.g., Andreasen, 2010; Šustek, 2011). However, exploring a high-dimensional parameter space usually also requires a considerable amount of likelihood evaluations. Hence, efficient techniques to evaluate the likelihood function become essential as the number of parameters increases.

In contrast to the frequentist approach, Bayesian econometricians treat the unknown parameters as random variables. By combining the information contained in the data (likelihood

[1] E.g., the European Central Bank uses an open-economy extension of the model by Smets and Wouters (2007), the so-called New Area-Wide Model, for macroeconomic projection exercises.

[2] Herbst (2015) reports that the likelihood is sometimes evaluated up to several million times, for both classical and the Bayesian approach.

function) with their prior beliefs about the parameters (prior density), the Bayesian approach seeks to gain information about the (posterior) density of the unknown parameters for a given set of data. From a technical perspective, modern sampling techniques, like the random walk Metropolis-Hasting (RWMH) algorithm used by Smets and Wouters (2007), the tailored randomized block Metropolis-Hastings (TaRBMH) algorithm suggested by Chib and Ramamurthy (2010), or the Sequential Monte Carlo sampler by Herbst and Schorfheide (2014), provide easily accessible ways to generate draws from the posterior distribution, that are often in some ways less challenging than maximizing the likelihood function. However, all three samplers mentioned above require a considerable amount of likelihood evaluations. For example, estimating a medium-scale DSGE model, such as the model by Smets and Wouters (2007), requires up to several million likelihood evaluations, depending on the selected sampling algorithm.[3] Thus not surprisingly, Herbst (2015) reports that, especially in medium to large-scale DSGEs models, the likelihood evaluation eventually becomes one of the bottlenecks in the estimation process.

As mentioned above, we can treat the policy function of (log-) linearized DSGE models as a linear SSM, where we assume that the behavior of a set of time series links to the dynamics of some potentially unobserved states. In the case of linear SSMs with Gaussian disturbances, we may use the so-called Kalman filter (KF) to recursively determine these states' mean vector and variance matrix for a given set of observed data. Consequently, the KF also provides the means to evaluate the likelihood function of the model. However, this recursive algorithm quickly becomes computationally demanding as the model's complexity increases. To reduce the computational burden of the KF, we might exploit the fact that for time-invariant and stationary SSMs, the uncertainty about the model's states converges towards an equilibrium as the number of observations increases. Hence, after a certain number of observations, it is no longer necessary to update the states' variance matrix, and we can replace the regular KF with a stationary recursion, which we refer to as a steady-state Kalman filter (SKF). However, the convergence process of the states' variance matrix can take many periods, especially when estimating DSGE models. Therefore, in this paper, we propose a variant of the KF based on a SKF augmented in the manner of de Jong (1988, 1991). We show that the additional computations caused by augmenting the filter require fewer arithmetic operations than those necessary to update the states' variance matrix. We find that this ASKF can significantly reduce the computational burden of the likelihood evaluation in medium- to large-scale SSMs. Furthermore, we show that for DSGE models without measurement error, where the number of exogenous state

[3]Note that Smets and Wouters (2007) generate only 250000 draws from the posterior distribution using the Random Walk Metropolis-Hastings sampler, where each draw is equivalent to one likelihood evaluation. However, as mentioned by Chib and Ramamurthy (2010), a careful exploration of the parameter space to find the mode of the posterior distribution needed to tune the Random Walk Metropolis-Hastings algorithm often requires a large amount of additional likelihood evaluations.

variables equals the number of the observable time series, it is usually possible to determine the equilibrium variance matrix of the model's states analytically.

The ASKF adds to a strand of literature that attempts to evaluate the likelihood of linear SSMs more efficiently. For models where the number of observed time series exceeds the number of states, Jungbacker and Koopman (2014) suggest a model transformation that reduces the dimensionality of the observation vector to the dimension of the state vector. This „collapsing of large observation vectors," as Durbin and Koopman (2012) call it, is particularly helpful in the context of dynamic factor models, where we attribute the common dynamics from a typically large number of time series to the movement in a small number of unobserved factors.

Koopman and Durbin (2000) propose a version of the KF in which they do not treat observations as vectors but consider each element of the observation vector as a new observation. Durbin and Koopman (2012, Chapter 6.4.4) show that this univariate treatment of multivariate observation vectors (UKF) requires fewer arithmetic operations than the regular KF, especially when the number of observed time series is large. Further, the UKF has proven particularly helpful when dealing with diffuse initialization problems.

Using the generic SSM of Chib and Ramamurthy (2010) – a simulation model with ten observable time series and five state variables – to compare the ASKF with the UKF, we find that the former eventually will outperform the latter, provided that the convergence process of the states' variance matrix lasts for at least 50 periods. Additionally, we use the model transformation by Jungbacker and Koopman (2014) to collapse the dimension of the observation vector, finding that, in this case, the ASKF becomes profitable after about 75 periods. The ASKF needs some periods to acclimatize because it requires determining the equilibrium variance matrix of the model's states prior to the actual recursion. However, since the convergence speed of the states' variance matrix typically cannot be determined ex-ante and the likelihood evaluation is usually relatively cheap in cases where the convergence process lasts only for a couple of periods, we consider the ASKF a valid option to evaluate the likelihood of SSMs, where the number of observed time series exceeds the number of states.

If, on the other hand, the number of states is significantly larger than the number of observable variables, as is often the case for structural DSGE models, the techniques mentioned above become less valuable. For this reason, Herbst (2015) suggests using the Chandrasekhar recursion (CR) developed by Morf (1974) and Morf et al. (1974) when estimating medium to large-scale DSGE models. Compared to the regular Kalman recursion, this algorithm replaces the Riccati difference equation (RDE), typically used to update the state variance matrix by another set of difference equations. When the number of state variables is significantly larger than the dimension of the observation vector, this set of "Chandrasekhar-type" difference equations can be shown to require fewer arithmetic operations than the original algorithm. We compare FORTRAN and MATLAB® implementations of the CR and the ASKF using the DSGE model introduced by Smets and Wouters (2007) as a benchmark. Even considering a variant of the Smets and Wouters (2007) model, in which all model variables are considered to be states, which is favorable for the CR, the FORTRAN implementation of the ASKF is almost twice as fast as the FORTRAN implementation of the CR. The ASKF performs even better in MATLAB®, being about three times quicker than the CR. Compared to the regular KF and the UKF, we find that the ASKF reduces the computational burden by 60 to 80 percent, depending on whether we consider all model variables as states or not.

The remainder of the paper reads as follows. The following section revisits some basic concepts necessary for the derivation of the ASKF. In Section 3, we will outline the basic idea of the ASKF and present an efficient algorithm to compute the log-likelihood of linear and time-invariant SSMs. Furthermore, we compare the additive and multiplicative operations of the regular KF and the ASKF for each additional observation and discuss the latter's implementation. In the subsequent section, we apply the ASKF to the generic SSM by Chib and Ramamurthy (2010) and the DSGE model introduced by Smets and Wouters (2007) and compare it in terms of speed and accuracy to the regular KF, the UKF, and the CR. The last section concludes the paper.

2 State-space models and the Kalman filter

In the following, we revisit some basic concepts and tools relevant throughout this paper. First, we introduce the class of linear and time-invariant state-space models (SSMs) and a textbook version of the Kalman filter (KF). Then we analyze the asymptotic properties of the filter and the concept of a steady-state Kalman filter (SKF). Finally, we introduce the augmented Kalman filter (AKF), which will form the foundation for deriving the augmented steady-state Kalman filter (ASKF) in the subsequent section.

2.1 State-space representation

As a general framework for our analysis, let us assume we have the following time-invariant, linear, and Gaussian SSM:[4]

$$\mathbf{y}_t = \mathbf{h} + \mathbf{H}\cdot\mathbf{w}_t + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \mathrm{N}(\mathbf{0},\mathbf{R}), \qquad \forall t = 1,2,\ldots,N \tag{1a}$$

$$\mathbf{w}_t = \mathbf{F}\cdot\mathbf{w}_{t-1} + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathrm{N}(\mathbf{0},\mathbf{Q}), \qquad \mathbf{w}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0,\mathbf{C}_0), \qquad \forall t = 1,2,\ldots,N \tag{1b}$$

where $\mathbf{y}_t \in \mathbb{R}^{n_y}$ and $\mathbf{w}_t \in \mathbb{R}^{n_w}$ are vectors containing the observed data and the potentially unobserved states at time t. The system matrices $\mathbf{F} \in \mathbb{R}^{n_w \times n_w}$, $\mathbf{H} \in \mathbb{R}^{n_y \times n_w}$, $\mathbf{Q} \in \mathbb{R}^{n_w \times n_w}$, $\mathbf{R} \in$

[4] We refer to Durbin and Koopman (2012, Chapter 4) for a textbook treatment of the KF with respect to SSMs, where the system matrices are allowed to change over time. However, note that most (log-) linearized DSGE and a variety of other time series models can be represented in terms of a time-invariant SSM.

$\mathbb{R}^{n_y \times n_y}$, and the vector $\mathbf{h} \in \mathbb{R}^{n_y}$ may be functions of a potentially uncertain vector $\boldsymbol{\theta}$ of time-invariant deep parameters. The normally distributed disturbances $\mathbf{u}_t \in \mathbb{R}^{n_y}$ and $\mathbf{v}_t \in \mathbb{R}^{n_w}$ are assumed to be serially independent and uncorrelated with each other, i.e.,

$$\mathrm{E}[\mathbf{u}_i \mathbf{u}_j^T] = \begin{cases} \mathbf{R}, & i = j, \\ \mathbf{0}, & i \neq j. \end{cases}, \quad \mathrm{E}[\mathbf{v}_i \mathbf{v}_j^T] = \begin{cases} \mathbf{Q}, & i = j, \\ \mathbf{0}, & i \neq j. \end{cases}, \quad \mathrm{E}[\mathbf{u}_i \mathbf{v}_j^T] = \mathbf{0}, \quad \forall\ i,\ j = 1, 2, \ldots, N.$$

Furthermore, they shall be uncorrelated to the initial state vector $\mathbf{w}_0$, so that

$$\mathrm{E}[\mathbf{u}_t(\mathbf{w}_0 - \boldsymbol{\mu}_\mathbf{0})^T] = \mathrm{E}[\mathbf{v}_t(\mathbf{w}_0 - \boldsymbol{\mu}_\mathbf{0})^T] = \mathbf{0}, \qquad \forall t = 1, 2, \ldots, N.$$

If all eigenvalues of the matrix $\mathbf{F}$ lie within the unit circle, we will call (1) a stationary SSM. Throughout this paper, we will use the SSM (1) as a flexible and general framework for the derivation and analysis of the ASKF. However, in some situations it will be convenient to consider the special case of the SSM (1) without the measurement error $\mathbf{u}_t$, i.e., $\mathbf{u}_t = \mathbf{0}$, $\forall t = 1, 2, \ldots, N$ and $\mathbf{R} = \mathbf{0}$, resulting in

$$\mathbf{y}_t = \mathbf{h} + \mathbf{H} \cdot \mathbf{w}_t, \qquad \forall t = 1, 2, \ldots, N \tag{2a}$$

$$\mathbf{w}_t = \mathbf{F} \cdot \mathbf{w}_{t-1} + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}), \qquad \mathbf{w}_0 \sim \mathrm{N}(\boldsymbol{\mu}_\mathbf{0}, \mathbf{C}_0), \qquad \forall t = 1, 2, \ldots, N \tag{2b}$$

Further, we will assume the system matrices $\mathbf{H}$, $\mathbf{F}$, and $\mathbf{Q}$ of SSM (2) to take the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_z & \mathbf{H}_x \end{pmatrix}, \tag{2c}$$

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_z \\ \mathbf{F}_x \end{pmatrix}, \tag{2d}$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{2e}$$

and that the quantities $\mathbf{w}_t$ and $\mathbf{v}_t$ define as

$$\mathbf{w}_t := \begin{pmatrix} \mathbf{z}_t^T & \mathbf{x}_t^T \end{pmatrix}^T, \qquad \mathbf{v}_t := \begin{pmatrix} \mathbf{v}_{t,z}^T & \mathbf{0} \end{pmatrix}^T, \qquad \forall\ t = 1, 2, \ldots, N,$$

where $\mathbf{x}_t \in \mathbb{R}^{n_x}$ represents the vector of the predetermined states already known from period $t-1$. In contrast, the vector $\mathbf{z}_t \in \mathbb{R}^{n_z}$ collects the exogenous states of the model, whose realization is affected by the stochastic innovations vector $\mathbf{v}_{t,z} \in \mathbb{R}^{n_z}$. The framework described by SSM (2) will meet the design requirements of a large number (log-) linearized DSGE models, such as the model introduced by Smets and Wouters (2007).

2.2 The Kalman filter

Both Bayesian and frequentist estimation techniques often require the evaluation of the likelihood function. A suitable tool for this purpose is the KF. Since the SSM (1) is Gaussian, for a given initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$ and a set of observations $\mathbf{Y}_N = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$ generated by SSM (1), the KF represents a recursive algorithm to compute the mean vectors — $\mathbf{w}_{t|t-1} := \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_{t-1}]$ and $\boldsymbol{\mu}_t := \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_t]$ — and variance matrices — $\mathbf{P}_{t|t-1} := \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_{t-1}]$ and $\mathbf{C}_t := \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_t]$ — of $\mathbf{w}_t$ given $\mathbf{Y}_{t-1}$ and $\mathbf{w}_t$ given $\mathbf{Y}_t$, respectively, for all periods $t = 1, 2, \ldots, N$. If we define

$$\mathbf{K}_t := \mathbf{P}_{t|t-1}\mathbf{H}^T \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}\right]^{-1}, \qquad \forall t = 1, 2, \ldots, N, \tag{3}$$

and let $\mathbf{e}_t := \mathbf{y}_t - \mathbf{h} - \mathbf{H}\mathbf{w}_{t|t-1}$ and $\mathbf{U}_t := \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}$ denote the forecast error of $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ and its corresponding variance matrix, respectively, we receive the Kalman recursion for $t = 1, 2, \ldots, N$ as

$$\begin{aligned}
\mathbf{w}_{t|t-1} &= \mathbf{F}\,\boldsymbol{\mu}_{t-1}, &\text{(4a)}\qquad \mathbf{P}_{t|t-1} &= \mathbf{F}\,\mathbf{C}_{t-1}\,\mathbf{F}^T + \mathbf{Q}, &\text{(4b)}\\
\mathbf{e}_t &= \mathbf{y}_t^{(\mathbf{h})} - \mathbf{H}\,\mathbf{w}_{t|t-1}, &\text{(4c)}\qquad \mathbf{U}_t &= \mathbf{H}\,\mathbf{P}_{t|t-1}\,\mathbf{H}^T + \mathbf{R}, &\text{(4d)}\\
\boldsymbol{\mu}_t &= \mathbf{w}_{t|t-1} + \mathbf{K}_t\,\mathbf{e}_t, &\text{(4e)}\qquad \mathbf{C}_t &= \mathbf{P}_{t|t-1} - \mathbf{K}_t\,\mathbf{H}\,\mathbf{P}_{t|t-1}, &\text{(4f)}
\end{aligned}$$

where $\mathbf{y}_t^{(\mathbf{h})} := \mathbf{y}_t - \mathbf{h}$ for all $t = 1, 2, \ldots, N$. A detailed derivation of recursion (4) is provided in Appendix A. Throughout this paper we will follow Lütkepohl (2007) and refer to the matrix $\mathbf{K}_t$ as the so-called Kalman gain.[5]

To avoid confusion, note that alternatively to the initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$ the KF may also be initialized at $(\mathbf{w}_{1|0}, \mathbf{P}_{1|0})$ with $\mathbf{w}_1 \sim \mathrm{N}(\mathbf{w}_{1|0}, \mathbf{P}_{1|0})$. In this case the state equation (1b) for $t = 1$ becomes redundant so that the corresponding SSM reduces to

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{h} + \mathbf{H}\cdot\mathbf{w}_t + \mathbf{u}_t, & \mathbf{u}_t &\sim \mathrm{N}(\mathbf{0}, \mathbf{R}), & & \forall t = 1, 2, \ldots, N, & \text{(5a)}\\
\mathbf{w}_t &= \mathbf{F}\cdot\mathbf{w}_{t-1} + \mathbf{v}_t, & \mathbf{v}_t &\sim \mathrm{N}(\mathbf{0}, \mathbf{Q}), & \mathbf{w}_1|\mathbf{Y}_0 &\sim \mathrm{N}(\mathbf{w}_{1|0}, \mathbf{P}_{1|0}), & \forall t = 2, 3, \ldots, N. & \text{(5b)}
\end{aligned}$$

The Kalman recursion for the SSM (5) is identical to recursion (4), apart from the fact that in $t = 1$ steps (4a) and (4b) are redundant, since $\mathbf{w}_{1|0}$ and $\mathbf{P}_{1|0}$ are already known. Thus, if we choose $\mathbf{w}_{1|0} = \mathbf{F}\boldsymbol{\mu}_0$ and $\mathbf{P}_{1|0} = \mathbf{F}\mathbf{C}_0\mathbf{F}^T + \mathbf{Q}$, the quantities computed by the KF are the same as the ones computed for the SSM (1). Although the Kalman filter is often derived based on the alternative SSM (5) (see e.g., Hamilton (1994, pp. 372-408), Durbin and Koopman (2012) or DeJong and Dave (2011)), hereafter we will focus on the SSM (1) presented at the beginning

[5]Note that some authors, e.g., Hamilton (1994) and Durbin and Koopman (2012), define the Kalman gain as $\mathbf{K}_t := \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^T \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T\right]^{-1}$. In this case the Kalman gain defines the gain matrix with respect to $\mathbf{w}_{t+1}$ given $\mathbf{Y}_t$, while in the current paper it is treated as the gain matrix with respect to $\mathbf{w}_t$ given $\mathbf{Y}_t$.

of this subsection, since it is more convenient for the derivation of the ASKF in Section 3.

As an important by-product the KF provides a possibility to evaluate the likelihood function of the SSM (1) for a given set of parameters $\boldsymbol{\theta}$ and a given initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$. To see this, note that $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ is normally distributed for all $t = 1, 2, \ldots, N$, with corresponding mean vector $\mathbf{h} + \mathbf{H}\mathbf{w}_{t|t-1}$ and variance matrix $\mathbf{U}_t$. Hence using the forecast-error decomposition the log-density of $\mathbf{Y}_N$ yields:

$$\log\left(f_{\mathbf{Y}_N}\right) = -\frac{n_y\, N}{2}\, \log(2\pi) - \frac{1}{2}\sum_{t=1}^{N} \log|\mathbf{U}_t| - \frac{1}{2}\sum_{t=1}^{N} \mathbf{e}_t^T\, \mathbf{U}_t^{-1}\, \mathbf{e}_t. \tag{6}$$

It is well-known (see e.g., Durbin and Koopman (2012, pp. 185)) that under quite general regularity conditions the distribution of the maximum-likelihood estimator for the deep parameters $\boldsymbol{\theta}$, defined by

$$\hat{\boldsymbol{\theta}} := \operatorname*{argmax}_{\boldsymbol{\theta}} \log\left(f_{\mathbf{Y}_N}\right),$$

is for large N approximately normally distributed with mean vector $\hat{\boldsymbol{\theta}}$ and variance matrix

$$\widehat{\mathrm{Var}}\left[\hat{\boldsymbol{\theta}}\right] = \left[\frac{\partial \log\left(f_{\mathbf{Y}_N}\right)}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}\right]^{-1}.$$

For a more detailed treatment of the asymptotic properties of the maximum-likelihood estimator, see e.g Hamilton (1994) and Harvey (1990a).

Taking the Bayesian perspective the density $f_{\mathbf{Y}_N}$ for a given parameter vector $\boldsymbol{\theta}$ is also important, since it is required to generate draws from posterior distribution $f_{\boldsymbol{\theta}|\mathbf{Y}_N} \propto f_{\mathbf{Y}_N|\boldsymbol{\theta}} \cdot f_{\boldsymbol{\theta}}$.

Note that the quantities of the KF, and therefore the log-likelihood defined by (6), are conditional on the distribution of the initial state vector $\mathbf{w}_0$, which itself is determined by $(\boldsymbol{\mu}_0, \mathbf{C}_0)$. The probably most common initialization strategy for stationary SSMs, is to specify $\boldsymbol{\mu}_0$ and $\mathbf{C}_0$ as the unconditional mean vector $\boldsymbol{\mu}$ and the unconditional variance matrix $\mathbf{C}$ of the state vector $\mathbf{w}_t$ (see e.g., Hamilton (1994, pp. 378) or Durbin and Koopman (2012, pp. 123,137-138)). These are obtainable from the state equation (1b) as

$$\boldsymbol{\mu} = \mathbf{0} \tag{7a}$$

and as the positive semi-definite matrix $\mathbf{C}$ solving the discrete Lyapunov equation

$$\mathbf{0} = \mathbf{F}\mathbf{C}\mathbf{F}^T + \mathbf{Q} - \mathbf{C}. \tag{7b}$$

This means that $\boldsymbol{\mu}_0$ and $\mathbf{C}_0$ are determined by $\mathbf{F}$ and $\mathbf{Q}$, which in turn are determined by the vector of deep parameters $\boldsymbol{\theta}$. Consequently, using this initialization (6) represents the exact or unconditional log-likelihood of the model.

2.3 The steady-state Kalman filter

Within the class of time-invariant and linear SSMs, it is a well-known feature of the KF that under certain circumstances the sequences $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$ converge towards fixed matrices. In this case we call the KF asymptotically time-invariant. To establish conditions for an asymptotically time-invariant filter, first note that in a time-invariant SSM like (1), the sequences $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$, obtained by the Kalman recursion (4) do not depend on the data itself. This becomes obvious if we use (4b), (4d), and (4f) to obtain the law of motion of the sequence $\{\mathbf{C}_t\}_{t=1}^N$ as

$$\mathbf{C}_t = \mathbf{F}\mathbf{C}_{t-1}\mathbf{F}^T + \mathbf{Q} - \left(\mathbf{F}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\mathbf{C}_{t-1}\mathbf{F}^T + \bar{\mathbf{G}}^T\right), \tag{8a}$$

with $\bar{\mathbf{H}} := \mathbf{H}\mathbf{F}$, $\bar{\mathbf{G}} := \mathbf{Q}\mathbf{H}^T$ and $\bar{\mathbf{R}} := \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R}$. Analogously, we can also obtain the law of motion of the sequence $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$ as

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^T - \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^T \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{F}^T + \mathbf{Q}. \tag{8b}$$

Furthermore, both (8a) and (8b) belong to the class of Riccati difference equations (RDEs), which convergence properties have been intensively studied by the literature (see e.g., Caines and Mayne (1970), Chan et al. (1984), de Souza et al. (1986) or De Nicolao and Gevers (1992)). In what follows, we give a brief summary of their results by establishing some well-known sufficient conditions under which the sequences $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$ converge against fixed matrices.[6]

We shall introduce some basic notions in advance: First, we call non-negative definite matrices $\mathbf{C}_+$ and $\mathbf{P}_+$ solutions of RDEs (8a) and (8b), respectively, if they satisfy the discrete algebraic Riccati equations (DAREs)

$$\mathbf{C}_+ = \mathbf{F}\mathbf{C}_+\mathbf{F}^T + \mathbf{Q} - \left(\mathbf{F}\mathbf{C}_+\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\mathbf{C}_+\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\mathbf{C}_+\mathbf{F}^T + \bar{\mathbf{G}}^T\right), \tag{9a}$$
$$\mathbf{P}_+ = \mathbf{F}\mathbf{P}_+\mathbf{F}^T - \mathbf{F}\mathbf{P}_+\mathbf{H}^T \left[\mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{P}_+\mathbf{F}^T + \mathbf{Q}, \tag{9b}$$

corresponding to (8a) and (8b), respectively. Furthermore, if $\mathbf{C}_+$ and $\mathbf{P}_+ = \mathbf{F}\mathbf{C}_+\mathbf{F}^T + \mathbf{Q}$ are solutions to RDE (8a) and (8b), respectively, we call them stabilizing / strong solutions, if and

[6] For a more general discussion on the convergence of time-invariant RDEs see Appendix B.

only if all eigenvalues of the matrix

$$\tilde{\mathbf{F}} = \mathbf{F}\left(\mathbf{I} - \mathbf{P}_+\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}\right]^{-1}\mathbf{H}\right) \tag{10}$$

are inside / inside or on the unit circle.[7] Consequently, every stabilizing solution is also a strong solution, while a strong solution is not necessarily also a stabilizing solution. Further, we may show that RDEs such as (8a) or (8b) have at most one (and therefore unique) strong solution (see e.g., de Souza et al., 1986).

Using this terminology we may establish three sufficient conditions under which the KF becomes asymptotically time-invariant:

Proposition 2.1 *Suppose $\mathbf{C}_0 \in \mathbb{R}^{n_w \times n_w}$ is an arbitrary, but symmetric and positive-definite matrix. Then in case of SSM (1) the sequences $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$ converge towards fixed matrices $\mathbf{C}_+$ and $\mathbf{P}_+$, i.e.*

$$\lim_{N\to\infty}\{\mathbf{C}_t\}_{t=1}^N = \mathbf{C}_+ \qquad \text{and} \qquad \lim_{N\to\infty}\{\mathbf{P}_{t|t-1}\}_{t=1}^N = \mathbf{P}_+,$$

if at least one of the following statements is true:

(i) The matrix $\mathbf{R}$ is non-singular and all eigenvalues of the matrix $\mathbf{F}$ are inside the unit-circle. In this case $\mathbf{C}_+$ and $\mathbf{P}_+$ are stabilizing solutions of the RDEs (8a) and (8b), respectively.

(ii) The matrix $\bar{\mathbf{R}}$ is non-singular and all eigenvalues of the matrix $\bar{\mathbf{F}} = \mathbf{F} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}$ are inside the unit-circle. In this case $\mathbf{C}_+$ and $\mathbf{P}_+$ are stabilizing solutions of the RDEs (8a) and (8b), respectively.

(iii) The matrix $\mathbf{C}_0$ satisfies the discrete Lyapunov equation

$$\mathbf{0} = \mathbf{F}\mathbf{C}_0\mathbf{F}^T + \mathbf{Q} - \mathbf{C}_0$$

and the all eigenvalues of the matrix $\mathbf{F}$ are inside the unit-circle. In this case $\mathbf{C}_+$ and $\mathbf{P}_+$ are strong solutions of the RDEs (8a) and (8b), respectively. Moreover, we may state that the matrices $\mathbf{C}_0 - \mathbf{C}_+$ and $\mathbf{P}_{1|0} - \mathbf{P}_+$ are positive-semi-definite.

We postpone the formal proof of Proposition 2.1 to Appendix B. From statement (i) of Proposition 2.1 follows that for all SSMs relying on a stationary process for $\mathbf{w}_t$ the KF becomes asymptotically time-invariant if the variance matrix of the measurement error $\mathbf{u}_t$ is non-singular. These assumptions are often met, for example, in the context of dynamic factor models (see e.g., Stock and Watson, 2016). There are however situations where $\mathbf{R}$ might be singular, but $\bar{\mathbf{R}}$ has full

[7] Note that some authors, e.g., Bini et al. (2012); Chiang et al. (2010), use the term almost stabilizing solution as synomym for a strong solution.

rank. This is often the case in the DSGE context (see e.g., Chari et al., 2007), as measurement errors are often omitted in these models. In this case we may use statement (ii) of Proposition 2.1, to investigate the convergence of (8a). If neither the conditions of statement (i) nor the conditions of statement (ii) are satisfied, statement (iii) will at least ensure that for any stationary SSM, both the matrix sequences, $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$, converge to an equilibrium, provided they are initialized at the unconditional variance matrix $\mathbf{C}$ of the state vector.

As mentioned before, all three statements of Proposition 2.1 provide sufficient conditions for the convergence of the matrix sequences $\{\mathbf{C}_t\}_{t=1}^N$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$. However, statements (i) and (ii) of Proposition 2.1 imply that $\mathbf{C}_+$ and $\mathbf{P}_+$ are stabilizing solutions of the RDEs (8a) and (8b), while statement (iii) guarantees only convergence to a strong solution. Note that convergence to a stabilizing solution has the major advantage that we may use standard methods, such as the Schur algorithm (Bini et al., 2012, Chapter 3.2), the Newton algorithm (Bini et al., 2012, Chapter 3.3) or the doubling algorithm (Anderson and Moore, 1979, Chapter 6.7), to numerically solve (9a) for its stabilizing solution. Although there are iterative algorithms for determining a strong solution, such as the structured doubling algorithm described by Bini et al. (2012, Chapter 5), these algorithms are potentially less efficient from a computational point of view (see e.g., Chiang et al., 2010).

One of the advantages of an asymptotically time-invariant filter is that at a certain period τ, when $\mathbf{C}_\tau$ has converged sufficiently close to $\mathbf{C}_+$, i.e., $\mathbf{C}_\tau \approx \mathbf{C}_{\tau-1}$, the Kalman recursion (4) for $\boldsymbol{\mu}_t$ might be replaced by

$$\boldsymbol{\mu}_{t,+} = \mathbf{K}_+ \, \mathbf{y}_t^{(\mathbf{h})} + \mathbf{J}_+ \, \boldsymbol{\mu}_{t-1,+}, \qquad \forall \; t = \tau+1, \tau+2, \ldots, N, \tag{11a}$$

with

$$\mathbf{P}_+ = \mathbf{F}\mathbf{C}_+\mathbf{F}^T + \mathbf{Q}, \tag{11b}$$
$$\mathbf{U}_+ = \mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}, \tag{11c}$$
$$\mathbf{K}_+ = \mathbf{P}_+\mathbf{H}^T \, \mathbf{U}_+^{-1}, \tag{11d}$$
$$\mathbf{J}_+ = (\mathbf{I} - \mathbf{K}_+\mathbf{H})\,\mathbf{F}. \tag{11e}$$

This usually reduces the computational burden significantly, since (11) does not involve the computationally expensive steps (4b), (4d) and (4f) of the original recursion. Furthermore, from (4a) and (4c) follows that the quantities $\mathbf{e}_t$ and $\mathbf{w}_{t|t-1}$ for $t = \tau+1, \tau+2, \ldots, N$ may be received in vectorized form as

$$\begin{pmatrix} \mathbf{w}_{\tau+1|\tau,+} & \cdots & \mathbf{w}_{N|N-1,+} \end{pmatrix} = \mathbf{F} \begin{pmatrix} \boldsymbol{\mu}_{\tau,+} & \cdots & \boldsymbol{\mu}_{N-1,+} \end{pmatrix}, \tag{11f}$$
$$\begin{pmatrix} \mathbf{e}_{\tau+1,+} & \cdots & \mathbf{e}_{N,+} \end{pmatrix} = \begin{pmatrix} \mathbf{y}_{\tau+1}^{(\mathbf{h})} & \cdots & \mathbf{y}_N^{(\mathbf{h})} \end{pmatrix} - \bar{\mathbf{H}} \begin{pmatrix} \boldsymbol{\mu}_{\tau,+} & \cdots & \boldsymbol{\mu}_{N-1,+} \end{pmatrix}, \tag{11g}$$

Throughout this paper, we will refer to the time-invariant filter described by (11) as the SKF. A detailed derivation of (11) is given in Appendix C.

Note that Hansen and Sargent (2013, pp. 160) suggest to initialize the KF at the stabilizing solution of (8a) or (8b), respectively, i.e., $\boldsymbol{\mu}_0 = \boldsymbol{\mu}_{0,+}$ and $\mathbf{C}_0 = \mathbf{C}_+$.[8] This initialization is often used in the context of maximum-likelihood estimation, since despite the fact that the log-likelihood calculated on the basis of this initialization usually does not reflect the exact or unconditional log-likelihood, the determined maximum-likelihood estimator may (under certain preconditions, see e.g., Harvey, 1990b, pp. 119, 129) have the same large-sample properties as the unconditional maximum-likelihood estimator. Although, the maximum-likelihood estimators based on an initialization different from the exact or unconditional initialization is in generally less efficient, the *steady-state* initialization has the advantage that the resulting log-likelihood can be computed using the quicker SKF (11) with $\tau = 0$. Furthermore, given $\mathbf{C}_0 = \mathbf{C}_+$ we can rewrite (6) in a more compact way:

$$\log\left(f_{\mathrm{Y}_N}\right)_+ = -\frac{1}{2}\left[n_y\ N\ \log(2\pi)\ + N\ \log|\mathbf{U}_+| + \mathrm{tr}\left(\mathbf{e}_{1:N,+}^T\ \mathbf{U}_+^{-1}\ \mathbf{e}_{1:N,+}\right)\right], \tag{12}$$

where $\mathbf{e}_{1:N,+} := \begin{pmatrix}\mathbf{e}_{1,+} & \cdots & \mathbf{e}_{N,+}\end{pmatrix}$.

2.4 The augmented Kalman filter

There may be situations where we want to investigate how the initial state vector (or some components of it) affects the quantities obtained by the KF. For instance, consider a SSM with several non-stationary states. A typical approach to initialize such a model is to consider the non-stationary elements of state vector as diffuse, which means that their variance will tend towards infinity. In such situations, it might be worth considering another variant of the KF, which goes back to influential work by de Jong (1988, 1991). In what follows, we briefly describe a version of what Durbin and Koopman (2012) call the augmented Kalman filter (AKF).[9]

Let us assume that $\bar{\mathbf{w}}_0 \in \mathbb{R}^{n_{\bar{w}}}$ and $\mathbf{d} \in \mathbb{R}^{n_d}$ with $n_{\bar{w}}, n_{\mathbf{d}} \leq n_w$ are two independent random vectors, such that we may write the initial state vector as

$$\mathbf{w}_0 = \mathbf{a}_w + \mathbf{A}_w\ \bar{\mathbf{w}}_0 + \mathbf{A}_d\ \mathbf{d}, \qquad \bar{\mathbf{w}}_0 \sim \mathrm{N}(\bar{\boldsymbol{\mu}}_0, \bar{\mathbf{C}}_0), \qquad \mathbf{d} \sim \mathrm{N}(\boldsymbol{\delta}_0, \mathbf{D}_0), \tag{13}$$

where $\mathbf{a}_w \in \mathbb{R}^{n_w}$, $\mathbf{A}_w \in \mathbb{R}^{n_w \times n_{\bar{w}}}$ and $\mathbf{A}_d \in \mathbb{R}^{n_w \times n_p}$. Note that (13) implies that $\mathbf{w}_0$ has the mean vector $\boldsymbol{\mu}_0 = \mathbf{a}_w + \mathbf{A}_w \bar{\boldsymbol{\mu}}_0 + \mathbf{A}_d \boldsymbol{\delta}_0$ and the variance matrix $\mathbf{C}_0 = \mathbf{A}_w \bar{\mathbf{C}}_0 \mathbf{A}_w^T + \mathbf{A}_d \mathbf{D}_0 \mathbf{A}_d^T$. By choosing $\mathbf{a}_w$, $\mathbf{A}_w$ and $\mathbf{A}_d$ appropriately, we may use (13) to decompose $\mathbf{w}_0$ into multiple components

[8]Note that in this context $\boldsymbol{\mu}_{0,+}$ is usually chosen to be unconditional mean vector $\boldsymbol{\mu} = \mathbf{0}$.

[9]A similar treatments of the AKF with respect to the alternative state-space representation (5) are given by Durbin and Koopman (2012, Chapter 5.7).

of interest. Durbin and Koopman (2012), for instance, choose $\mathbf{a}_w$, $\mathbf{A}_\mathbf{w}$ and $\mathbf{A}_\mathbf{d}$, such that $\mathbf{a}_w$ represents the fixed (and observable) elements, $\bar{\mathbf{w}}_0$ the stationary elements, and $\mathbf{d}$ the diffuse elements of $\mathbf{w}_0$.

Now suppose, we specified $\mathbf{w}_0$ according to (13) and want to examine how the distribution of the random vector $\mathbf{d}$ affects the quantities $\boldsymbol{\mu}_t$ and $\mathbf{C}_t$ of the KF. As we show in Appendix D, denoting the time t quantities generated by the Kalman recursion (4) initialized at $(\tilde{\boldsymbol{\mu}}_0, \tilde{\mathbf{C}}_0)$, with

$$\tilde{\boldsymbol{\mu}}_0 = \mathbf{a}_w + \mathbf{A}_w\, \bar{\boldsymbol{\mu}}_0, \tag{14a}$$

$$\tilde{\mathbf{C}}_0 = \mathbf{A}_w\, \bar{\mathbf{C}}_0\, \mathbf{A}_w^T, \tag{14b}$$

by $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$ and $\tilde{\mathbf{K}}_t$, we may express $\boldsymbol{\mu}_t$ and $\mathbf{C}_t$ as

$$\boldsymbol{\mu}_t = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\mathbf{A}_d\, \left(\mathbf{D}_0^{-1} + \mathbf{A}_d^T\, \mathbf{S}_t\, \mathbf{A}_d\right)^{-1} \left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_d^T\, \mathbf{s}_t\right), \qquad \forall\ t = 1, 2, \ldots, N, \tag{15a}$$

$$\mathbf{C}_t = \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d \left(\mathbf{D}_0^{-1} + \mathbf{A}_d^T\, \mathbf{S}_t\, \mathbf{A}_d\right)^{-1} \mathbf{A}_d^T\mathbf{M}_t^T, \qquad \forall\ t = 1, 2, \ldots, N, \tag{15b}$$

where we may obtain $\mathbf{s}_t$, $\mathbf{S}_t$ and $\mathbf{M}_t$ recursively as

$$\mathbf{s}_t = \mathbf{s}_{t-1} + (\mathbf{HFM}_{t-1})^T\, \tilde{\mathbf{U}}_t^{-1}\, \tilde{\mathbf{e}}_t, \qquad \forall\ t = 1, 2, \ldots, N, \tag{15c}$$

$$\mathbf{S}_t = \mathbf{S}_{t-1} + (\mathbf{HFM}_{t-1})^T\, \tilde{\mathbf{U}}_t^{-1}\, (\mathbf{HFM}_{t-1}), \qquad \forall\ t = 1, 2, \ldots, N, \tag{15d}$$

$$\mathbf{M}_t = \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{F}\, \mathbf{M}_{t-1}, \qquad \forall\ t = 1, 2, \ldots, N, \tag{15e}$$

with $\mathbf{s}_0 = \mathbf{0}$, $\mathbf{S}_0 = \mathbf{0}$ and $\mathbf{M}_0 = \mathbf{I}$. Furthermore, we may write the log-density of $\mathbf{Y}_N$ as

$$\begin{aligned}\log(f_{\mathbf{Y}_N}) = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\log|\mathbf{I} + \mathbf{D}_0\, \mathbf{A}_d^T\, \mathbf{S}_N\, \mathbf{A}_d| - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_\mathbf{0}^{-1}\boldsymbol{\delta}_0 \\ + \frac{1}{2}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_d^T\, \mathbf{s}_N)^T(\mathbf{D}_0^{-1} + \mathbf{A}_d^T\, \mathbf{S}_N\, \mathbf{A}_d)^{-1}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_d^T\, \mathbf{s}_N),\end{aligned} \tag{16}$$

where $\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)$ represents the log-density of $\mathbf{Y}_N$ given $\mathbf{d} = \mathbf{0}$.[10] Thus, using the AKF (15) we can directly examine the effect of $\boldsymbol{\delta}_0$ and $\mathbf{D}_0$ on $\boldsymbol{\mu}_t$, $\mathbf{C}_t$ and $\log(f_{\mathbf{Y}_N})$ from (15a), (15b), and (16). This allows us to study the special cases in which parts of $\mathbf{w}_0$ are considered to be fixed (i.e., $\mathbf{D}_0 \to \mathbf{0}$) or diffuse (i.e., $\mathbf{D}_0 \to \infty$). In Appendix D we provide a digression on how to incorporate initialization strategies for non-stationary SSMs within the AKF (15).

[10] Note that the recursion for $\mathbf{s}_t$, $\mathbf{S}_t$ and $\mathbf{M}_t$ can be incorporated in the Kalman recursion used to compute $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$ and $\tilde{\mathbf{K}}_t$. A detailed derivation of equations (15)-(16) and how these steps can be incorporated in the Kalman recursion (4) is provided in Appendix D. A similar treatment of the AKF with respect to the alternative state-space representation (5) is given by Durbin and Koopman (2012, pp. 141-146).

3 Augmented steady-state Kalman filter

Equipped with the concepts introduced in the previous section, we now turn our attention to the derivation of the ASKF. For this purpose, we will first outline the basic idea of an ASKF and obtain a general algorithm to evaluate the likelihood function of a linear and time-invariant SSM. We show that compared to the standard KF this algorithm lowers the computational burden associated with each additional observation. Finally, we provide conditions for applicability of the ASKF which we show are satisfied for all stationary SSMs. Further, we show that for the SSM (2) with $n_y = n_z$ the algorithm can be additionally optimized, since in this case there is an analytical solution to RDE (8a).

3.1 Basic idea

Suppose we want to evaluate the log-density $\log(f_{\mathbf{Y}_N})$ of the SSM (1) for a given initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$. Furthermore, suppose that the RDE (8a) has a solution $\mathbf{C}_+$ such that $\mathbf{C}_0 - \mathbf{C}_+$ is a positive semi-definite matrix.

As mentioned earlier, we can determine the $\log(f_{\mathbf{Y}_N})$ for this initialization using the KF (4) and equation (6). However, we could also employ the AKF (15) along with equation (16) to determine $\log(f_{\mathbf{Y}_N})$. To do so, we need to specify the model (13) for the initial state vector $\mathbf{w}_0$ such that

$$\boldsymbol{\mu}_0 = \mathbf{a}_w + \mathbf{A}_w\, \bar{\boldsymbol{\mu}}_0 + \mathbf{A}_d\, \boldsymbol{\delta}_0, \tag{17a}$$

$$\mathbf{C}_0 = \mathbf{A}_w\, \bar{\mathbf{C}}_0\, \mathbf{A}_w^T + \mathbf{A}_d\, \mathbf{D}_0\, \mathbf{A}_d^T, \tag{17b}$$

ensuring that $\mathbf{w}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \mathbf{C}_0)$. There are several possible specifications of (13) which satisfy (17a) and (17b). The basic idea of the ASKF is to choose the model (13) for the initial state vector $\mathbf{w}_0$ in a way that makes the Kalman recursion, on which the AKF is based, time-invariant. To do so, we will specify (13) as follows: First, we will set

$$\mathbf{a}_w = \mathbf{0}, \tag{18a}$$

$$\mathbf{A}_w = \mathbf{I}, \tag{18b}$$

$$\bar{\boldsymbol{\mu}}_0 = \boldsymbol{\mu}_0, \tag{18c}$$

$$\bar{\mathbf{C}}_0 = \mathbf{C}_+, \tag{18d}$$

so that from (14a) and (14b) follows $\tilde{\boldsymbol{\mu}}_0 = \boldsymbol{\mu}_0$ and $\tilde{\mathbf{C}}_0 = \mathbf{C}_+$, respectively. Second, in order to satisfy (17a) and (17b), we set

$$\boldsymbol{\delta}_0 = \mathbf{0}, \tag{18e}$$

$$\mathbf{D}_0 = \mathbf{I}, \tag{18f}$$

and choose $\mathbf{A}_d$ such that

$$\mathbf{A}_d \mathbf{A}_d^T = \mathbf{C}_0 - \mathbf{C}_+. \tag{18g}$$

Hence, the fully specified version of (13) yields

$$\mathbf{w}_0 = \bar{\mathbf{w}}_0 + \mathbf{A_d}\ \mathbf{d}, \qquad \bar{\mathbf{w}}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \mathbf{C}_+), \qquad \mathbf{d} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}). \tag{19}$$

In the current paper, we use the singular value decomposition of $\mathbf{C}_0 - \mathbf{C}_+$ to obtain $\mathbf{A}_d$. However, note that in the cases where $\mathbf{C}_0 - \mathbf{C}_+$ is non-singular, one might also set $\mathbf{A}_d = \mathbf{I}$ and $\mathbf{D}_0 = \mathbf{C}_0 - \mathbf{C}_+$. In either way, we need to ensure that $\mathbf{D}_0$ is positive definite.

Specifying (13) this way, the quantities $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$ and $\tilde{\mathbf{K}}_t$ corresponding to the KF (4) initialized at $(\tilde{\boldsymbol{\mu}}_0, \tilde{\mathbf{C}}_0)$ become

$$\begin{aligned} &\tilde{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_{t,+}, && \tilde{\mathbf{w}}_{t|t-1} = \mathbf{w}_{t|t-1,+}, && \tilde{\mathbf{e}}_t = \mathbf{e}_{t,+}, && \\ &\tilde{\mathbf{C}}_t = \mathbf{C}_+, && \tilde{\mathbf{P}}_{t|t-1} = \mathbf{P}_+, && \tilde{\mathbf{U}}_t = \mathbf{U}_+, && \tilde{\mathbf{K}}_t = \mathbf{K}_+, \quad \forall\ t = 1, 2, \ldots, N, \end{aligned}$$

where $\boldsymbol{\mu}_{t,+}$, $\mathbf{C}_+$, $\mathbf{w}_{t|t-1,+}$, $\mathbf{P}_+$, $\mathbf{e}_{t,+}$, $\mathbf{U}_+$ and $\mathbf{K}_+$ are the quantities computed by the SKF based on the initialization $(\boldsymbol{\mu}_{0,+}, \mathbf{C}_+)$, with $\boldsymbol{\mu}_{0,+} = \boldsymbol{\mu}_0$. This reduces the computational burden of the AKF in two ways: First, instead of the regular KF we may use the faster SKF to compute the quantities $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$, $\tilde{\mathbf{K}}_t$ and $\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)$. Second, we can simplify the recursion (15c)-(15e), since the expression $\left(\mathbf{I} - \tilde{\mathbf{K}}_t \mathbf{H}\right)\mathbf{F}$ becomes time-invariant and identically to $\mathbf{J}_+$. Thus, the quantity $\mathbf{M}_t$ reduces to

$$\mathbf{M}_t = \mathbf{J}_+^t, \qquad \forall\ t = 0, 1, \ldots, N. \tag{20}$$

3.2 Likelihood evaluation

Furthermore, suppose our interest lies exclusively in the evaluation of the log-density $\log\left(f_{\mathbf{Y}_N}\right)$. If we specify (13) according to (18), we can simplify (16) to

$$\log(f_{\mathbf{Y}_N}) = \log\left(f_{\mathbf{Y}_N}\right)_+ - \frac{1}{2}\log|\mathbf{I} + \mathbf{A}_d^T\ \mathbf{S}_N\ \mathbf{A}_d| + \frac{1}{2}\mathbf{s}_N^T\ \mathbf{A}_d\ (\mathbf{I} + \mathbf{A}_d^T\ \mathbf{S}_N\ \mathbf{A}_d)^{-1}\ \mathbf{A}_d^T\ \mathbf{s}_N, \tag{21}$$

where $\log\left(f_{\mathbf{Y}_N}\right)_+ = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)$ represents the log-density obtained from the SKF based on the initialization $(\boldsymbol{\mu}_{0,+}, \mathbf{C}_+)$, with $\boldsymbol{\mu}_{0,+} = \boldsymbol{\mu}_0$. Hence, the log-density $\log\left(f_{\mathbf{Y}_N}\right)$ is fully determined by $\log\left(f_{\mathbf{Y}_N}\right)_+$, $\mathbf{s}_N$, $\mathbf{S}_N$, and $\mathbf{A}_d$.

It turns out that we may further optimize the computation of $\mathbf{s}_N$, $\mathbf{S}_N$, and $\log\left(f_{\mathbf{Y}_N}\right)_+$, in terms of the required arithmetic operations. To do so, let us define

$$\mathbf{b}_t := \mathbf{V}^T \mathbf{e}_{t,+}, \qquad \forall\ t = 1, 2, \ldots, N, \tag{22a}$$

$$\mathbf{B}_t := \left(\mathbf{J}_+^t\right)^T \bar{\mathbf{H}}^T\ \mathbf{V}, \qquad \forall\ t = 0, 1, \ldots, N-1, \tag{22b}$$

with $\mathbf{V}$ satisfying $\mathbf{U}_+^{-1} = \mathbf{V}\mathbf{V}^T$, so that we may obtain $\mathbf{s}_N$ and $\mathbf{S}_N$ as

$$\mathbf{s}_N = \mathbf{B}_{0:N-1}\ \operatorname{vec}(\mathbf{b}_{1:N}), \tag{23a}$$

$$\mathbf{S}_N = \mathbf{B}_{0:N-1}\ \mathbf{B}_{0:N-1}^T, \tag{23b}$$

with $\mathbf{B}_{0:N-1} := \left(\mathbf{B}_0, \cdots, \mathbf{B}_{N-1}\right)$ and $\mathbf{b}_{1:N} := \left(\mathbf{b}_1, \cdots, \mathbf{b}_N\right)$. It follows from (22) that $\mathbf{b}_{1:N}$ yields

$$\mathbf{b}_{1:N} = \mathbf{V}^T \mathbf{e}_{1:N,+}, \tag{24}$$

and that $\mathbf{B}_{0:N-1}$ is recursively defined by

$$\mathbf{B}_t := \mathbf{J}_+^T\ \mathbf{B}_{t-1} \qquad \forall\ t = 1, 2, \ldots, N-1, \tag{25}$$

with $\mathbf{B}_0 = \bar{\mathbf{H}}^T\ \mathbf{V}$. Hence, using (24) to rewrite (12) as

$$\log\left(f_{\mathbf{Y}_N}\right)_+ = -\frac{1}{2}\left[n_y\ N\ \log(2\pi)\ + N\ \log|\mathbf{U}_+| + \operatorname{tr}\left(\mathbf{b}_{1:N}\ \mathbf{b}_{1:N}^T\right)\right], \tag{26}$$

the log-density $\log\left(f_{\mathbf{Y}_N}\right)_+$ is obtainable by performing the steps displayed in Algorithm 1.

The ASKF described in Algorithm 1 has several advantages compared to the regular KF, where the probably most important is that, although the initial setup (Steps (1)-(3)) of the filter is more expansive, it requires fewer arithmetic operations for each additional observation, so that the recursive part of the algorithm is more efficient. To see this, suppose we implement the KF and the ASKF using standard matrix multiplication, ignoring the advantages which may arise from the symmetrical nature of variance matrices. For this case, Table 1 lists the number of additional additive and multiplicative operations that result if we increase the number of observations from N to $N+1$. It turns out that the ASKF saves $n_y^2 + n_w(n_w^2 - 1) + n_w(n_w - 1) + n_w(n_w - 1)^2 + 2n_y(n_y - 1)(n_w - 1)$ additive and $2n_w^3 + n_y + n_y^2(n_w - 1) + n_y(n_y - 1)n_w$ multiplicative operations for each additional observation. Further, and almost more importantly, the ASKF does not require the repeated computation of the inverse and the determinant of the $n_y \times n_y$

Algorithm 1: Using the ASKF to compute $\log\left(f_{\mathbf{Y}_N}\right)$ for the SSM (1)

Input : μ_0, $\mathbf{C}_0$ and $\mathbf{C}_+$, where $\mathbf{C}_0 - \mathbf{C}_+$ is positive semi-definite;

Execute: Steps (1)-(7);

(1) Choose $\mathbf{a}_w$, $\mathbf{A}_w$, $\bar{\mu}_0$, $\bar{\mathbf{C}}_0$, δ_0, $\mathbf{D}_0$, and $\mathbf{A}_d$ according to (18);

(2) Obtain $\mathbf{P}_+$, $\mathbf{U}_+$, $\mathbf{K}_+$ and $\mathbf{J}_+$ from (11b)-(11e);

(3) Obtain $\mathbf{V}$ such that $\mathbf{U}_+^{-1} = \mathbf{V}\mathbf{V}^T$ and set $\mathbf{B}_0 = \bar{\mathbf{H}}^T\,\mathbf{V}$;

(4) **for** $t = 1$ **to** $N-1$ **do**

Obtain $\mu_{t,+}$ (with $\mu_{0,+} = \mu_0$) and $\mathbf{B}_t$ from (11a) and (25);

(5) Obtain $\mathbf{e}_{1:N,+}$ and $\mathbf{b}_{1:N}$ from (11g) and (24);

(6) Obtain $\mathbf{s}_N$, $\mathbf{S}_N$ and $\log\left(f_{\mathbf{Y}_N}\right)_+$ from (23a), (23b) and (26);

(7) Obtain $\log\left(f_{\mathbf{Y}_N}\right)$ from (21);

Output : $\log\left(f_{\mathbf{Y}_N}\right)$;

matrix $\mathbf{U}_t$. Finally, the ASKF offers more room for parallelization since only the computation of $\mu_{t,+}$ and $\mathbf{B}_t$ are strictly sequential.

3.3 Requirements to apply the augmented steady-state Kalman filter

To apply Algorithm 1 to the SSM (1) for a given initialization $(\mu_0, \mathbf{C}_0)$, we need to satisfy the following assumptions:

Assumption 3.1 *We assume that*

(i) there is a solution $\mathbf{C}_+$ *to the RDE* (8a)*,*

(ii) such that the matrix $\mathbf{C}_0 - \mathbf{C}_+$ *is positive semi-definite.*

While Assumption 3.1(i) is needed to use the SKF at all, Assumption 3.1(ii) is necessary to satisfy (17b), since $\mathbf{D}_0$ is positive-definite by definition, so that $\mathbf{A}_d\mathbf{D}_0\mathbf{A}_d^T$ must be at least positive-semi-definite. Although it is not necessary in theory, in practice, it is advisable to ensure that $\mathbf{C}_+$ represents a strong solution of RDE (8a). Otherwise the matrix $\mathbf{J}_+$ possesses explosive eigenvalues, making Algorithm 1 numerically unstable.[11]

[11] Note that we show in Appendix B and C that the matrices $\tilde{\mathbf{F}}$ and $\mathbf{J}_+$ share the same set of eigenvalues.

Table 1: Computational expanse of an additional observation

Eqn.	Additional Matrix Operations	Multiplications	Additions
	augmented steady-state Kalman filter		
Eqn.	Additional Matrix Operations	Multiplications	Additions
(11a)	$(n_w \times n_y)[(n_y \times 1)-(n_y \times 1)]+(n_w \times n_w)(n_w \times 1)$	$n_y n_w + n_w^2$	$n_y n_w + n_w^2 + n_y - n_w$
(25)	$(n_w \times n_w)(n_w \times n_y)$	$n_y n_w^2$	$n_y n_w^2 - n_y n_w$
(11g)	$(n_y \times 1)-(n_y \times n_w)(n_w \times 1)$	$n_y n_w$	$n_y n_w$
(24)	$(n_y \times n_y)(n_y \times 1)$	n_y^2	$n_y^2 - n_y$
(23a)	$(n_w \times 1)+(n_w \times n_y)(n_y \times 1)$	$n_y n_w$	$n_y n_w$
(23b)	$(n_w \times n_w)+(n_w \times n_y)(n_y \times n_w)$	$n_y n_w^2$	$n_y n_w^2$
(26)	$(n_y \times n_y)+(n_y \times 1)(1 \times n_y)$	n_y^2	n_y^2
m_{ASKF}	$2n_y^2 + n_w^2 + 2n_y n_w^2 + 3n_y n_w$		
a_{ASKF}	$2n_y^2 + n_w^2 - n_w + 2n_y n_w^2 + 2n_y n_w$		
	Kalman filter		
Eqn.	Additional Matrix Operations	Multiplications	Additions
(4a)	$(n_w \times n_w)(n_w \times 1)$	n_w^2	$n_w^2 - n_w$
(4b)	$(n_w \times n_w)(n_w \times n_w)(n_w \times n_w)+(n_w \times n_w)$	$2n_w^3$	$2n_w^3 - n_w^2$
(4c)	$(n_y \times 1)-(n_y \times 1)-(n_y \times n_w)(n_w \times 1)$	$n_y n_w$	$n_y n_w + 2n_y - n_w$
(4d)	$(n_y \times n_w)(n_w \times n_w)(n_w \times n_y)+(n_y \times n_y)$	$n_y n_w^2 + n_y^2 n_w$	$n_y n_w^2 + n_y^2 n_w - n_y n_w$
(3)	$(n_w \times n_y)(n_y \times n_y)$	$n_y^2 n_w$	$n_y^2 n_w - n_y n_w$
(4e)	$(n_w \times 1)+(n_w \times n_y)(n_y \times 1)$	$n_y n_w$	$n_y n_w$
(4f)	$(n_w \times n_w)-(n_w \times n_y)(n_y \times n_w)$	$n_y n_w^2$	$n_y n_w^2$
(6)	$(1 \times 1)+(1 \times n_y)(n_y \times n_y)(n_y \times 1)$	$n_y^2 + n_y$	n_y^2
m_{KF}	$2n_w^3 + n_y^2 + n_w^2 + n_y + 2n_y^2 n_w + 2n_y n_w^2 + 2n_y n_w$		
a_{KF}	$2n_w^3 + n_y^2 + 2n_y - 2n_w + 2n_y^2 n_w + 2n_y n_w^2$		
	Comparison		
$m_{KF} - m_{ASKF}$	$2n_w^3 + n_y + n_y^2(n_w - 1) + n_y(n_y - 1)n_w$		
$a_{KF} - a_{ASKF}$	$n_y^2 + n_w(n_w^2 - 1) + n_w(n_w - 1) + n_w(n_w - 1)^2 + 2n_y(n_y - 1)(n_w - 1)$		

We count $nm(l-1)$ additive and nml multiplicative operations for product of a $m \times l$ and a $l \times n$ matrix. Further, nm additive operations are counted for the sum/difference of two $m \times n$ matrices. The a_{KF} and a_{ASKF} denote the number additive operations of the corresponding filter, while m_{KF} and m_{ASKF} denote their required number of multiplicative operations.

Stationary state-space models: To check how restrictive the conditions of Assumption 3.1 are, let us first consider the class of stationary SSMs. Thus, we consider models where all the eigenvalues of the transition matrix $\mathbf{F}$ lie within the unit circle, or in other words where $\mathbf{F}$ is stable. Thus, from Corollary 2.1(iii), we know that for stationary SSMs, the preconditions for the usage of the ASKF will be satisfied, provided we use the unconditional variance of $\mathbf{w}_t$ to initialize the SSM (1), i.e., $\mathbf{C}_0 = \mathbf{C}$. To obtain $\mathbf{C}_+$, which in this case is a strong solution to (8a), we could use an iterative algorithm, such as the structured doubling algorithm described by Bini et al. (2012, Chapter 5). It is convenient to consider two special cases where $\mathbf{C}_+$ we may obtain in a different manner.

Case 1: If we consider a stationary SSM, where the variance matrix $\mathbf{R}$ of the measurement error is positive-definite, it follows from Corollary 2.1(i) that the RDE (8a) converges to a stabilizing solution $\mathbf{C}_+$ for any positive-definite initialization $\mathbf{C}_0$. Furthermore, $\mathbf{C}_+$ can be considered the only non-negative-definite solution of the RDE (8a).[12] Thus, using the unconditional ini-

[12]This follows from Proposition B.1(ii) in Appendix B.

tialization, i.e., $\mathbf{C}_0 = \mathbf{C}$, we may apply the classic techniques discussed in the previous section to obtain the stabilizing solution $\mathbf{C}_+$ of RDE (8a). The same argumentation applies to stationary SSMs, where $\bar{\mathbf{R}}$ is positive-definite and where $\bar{\mathbf{F}}$ is stable.

Case 2: Now, let us consider SSMs of the form (2), where the number of observable variables n_y equals the number of exogenous states variables n_z. In the following Proposition, we will show that in this case, we may obtain $\mathbf{C}_+ = \mathbf{0}$ as a solution to the RDE (8a) if the matrix $\mathbf{H}_z$ is non-singular.

Proposition 3.1 *Suppose there is a SSM of the form described by* (2) *where*

(i) the number of observable variables n_y equals the number of (state-) disturbances n_z,

(ii) and where the matrix $\mathbf{H}_z$ is non-singular.

Then $\mathbf{C}_+ := \mathbf{0}$ is a solution to the RDE (8a).

Proof:

To prove that $\mathbf{C}_+ := \mathbf{0}$ is a solution to the RDE (8a), it is sufficient to show that $\mathbf{C}_+ := \mathbf{0}$ satisfies the DARE

$$\mathbf{C}_+ = \bar{\mathbf{F}}\mathbf{C}_+\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\mathbf{C}_+\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\mathbf{C}_+\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\mathbf{C}_+\bar{\mathbf{F}}^T. \tag{27}$$

with $\bar{\mathbf{F}} := \mathbf{F} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}$ and $\bar{\mathbf{Q}} := \mathbf{Q} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T$, which, due to Lemma B.1 in Appendix B, is equivalent to the DARE (9a) corresponding to the RDE (8a). To see that $\mathbf{C}_+ := \mathbf{0}$ is a solution to (27), note that for the SSM (2)

$$\bar{\mathbf{R}} = \mathbf{H}\mathbf{Q}\mathbf{H}^T = \begin{pmatrix}\mathbf{H}_z & \mathbf{H}_x\end{pmatrix}\begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{H}_z^T\\ \mathbf{H}_x^T\end{pmatrix} = \begin{pmatrix}\mathbf{H}_z\mathbf{Q}_z & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{H}_z^T\\ \mathbf{H}_x^T\end{pmatrix} = \mathbf{H}_z\mathbf{Q}_z\mathbf{H}_z^T.$$

Thus, from the definition of $\bar{\mathbf{Q}}$ follows that

$$\begin{aligned}
\bar{\mathbf{Q}} &= \mathbf{Q} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T = \mathbf{Q} - \mathbf{Q}\mathbf{H}^T\left(\mathbf{H}_z\mathbf{Q}_z\mathbf{H}_z^T\right)^{-1}\mathbf{H}\mathbf{Q}\\
&= \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{H}_z^T\\ \mathbf{H}_x^T\end{pmatrix}(\mathbf{H}_z\,\mathbf{Q}_z\,\mathbf{H}_z^T)^{-1}\begin{pmatrix}\mathbf{H}_z & \mathbf{H}_x\end{pmatrix}\begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \begin{pmatrix}\mathbf{Q}_z\mathbf{H}_z^T\\ \mathbf{0}\end{pmatrix}(\mathbf{H}_z^T)^{-1}\,\mathbf{Q}_z^{-1}\,\mathbf{H}_z^{-1}\begin{pmatrix}\mathbf{H}_z\mathbf{Q}_z & \mathbf{0}\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \begin{pmatrix}\mathbf{Q}_z\\ \mathbf{0}\end{pmatrix}\mathbf{Q}_z^{-1}\begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\end{pmatrix}\\
&= \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} - \begin{pmatrix}\mathbf{Q}_z & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix} = \mathbf{0}.
\end{aligned}$$

This, however, means that (27) simplifies to

$$\mathbf{C}_+ = \bar{\mathbf{F}}\mathbf{C}_+\bar{\mathbf{F}}^T - \bar{\mathbf{F}}\mathbf{C}_+\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\mathbf{C}_+\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\mathbf{C}_+\bar{\mathbf{F}}^T,$$

which is clearly satisfied if we set $\mathbf{C}_+ = \mathbf{0}$.

□

This result is convenient since $\mathbf{C}_+ = \mathbf{0}$ satisfies Assumption 3.1(ii) for any positive semi-definite $\mathbf{C}_0$.[13] This means for $n_y = n_z = \text{rk}(\mathbf{H}_z)$, we do not even have to obtain $\mathbf{C}_+$ numerically and may use Algorithm 1 to compute $\log\left(f_{\mathbf{Y}_N}\right)$ for an arbitrary initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$.

At this point, it is appropriate to discuss why most DSGE models without measurement errors will meet the preconditions of Proposition 3.1. To see this, first, consider the scenario where $n_y > n_z$. This, in general, will lead to a model that is unable to match the data.[14] Consequently, we will have to include an appropriate number of measurement disturbances into our SSM. However, adding zeros to the corresponding entries in the transitions matrix $\mathbf{F}$ we may treat these measurement disturbances as state disturbances and get yet a SSM without measurement error satisfying condition (i) of Proposition 3.1. Thus $n_y > n_z$ is less of a problem. Now consider the opposite case where $n_y < n_z$, which usually implies that our model can replicate the data with more than one set of state disturbances. Admittedly this does not necessarily pose a problem in estimating the model using the common KF (4). However, in this case, it is often possible to include additional observable data series to increase the information used to estimate the model's parameters. This is especially true in the DSGE context since the observation vector $\mathbf{y}_t$ often reflects only a fraction of the potentially observable variables for these models.

Assuming $n_z = n_y$ holds, we will usually find that condition (ii) of Proposition 3.1 is also satisfied. To see this, note that there are only two possible scenarios where the model might not satisfy condition (ii) of Proposition 3.1: In the first scenario, we have the trivial case with $\text{rk}(\mathbf{H}_z) \leq \text{rk}(\mathbf{H}) < n_y$, indicating that a fraction of the observations vector $\mathbf{y}_t$ can be written as a linear combination of the remaining set of observations in $\mathbf{y}_t$ and thus contains redundant information. More interesting is the second scenario, where $\text{rk}(\mathbf{H}_z) < \text{rk}(\mathbf{H}) = n_y$. In this case, a fraction of $\mathbf{y}_t$ contains information that, from the model's perspective, was already determined in the previous period $t-1$. To see this, note that we can reorder the observations vector $\mathbf{y}_t$ so that we can write

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_x^{(1)} \\ \mathbf{H}_x^{(2)} \end{bmatrix} \mathbf{x}_t + \begin{bmatrix} \mathbf{H}_z^{(1)} \\ \mathbf{H}_z^{(2)} \end{bmatrix} \mathbf{z}_t \tag{28}$$

[13] As mentioned before, in practice, we might have to check if $\mathbf{C}_+ = \mathbf{0}$ represents a strong solution.
[14] We exclude the trivial case with redundant observations that are linear combinations of the remaining set of data.

with $\text{rk}(\mathbf{H}_z^{(1)}) = \text{rk}(\mathbf{H}_z)$. This, however, means there is a matrix $\boldsymbol{\Gamma}$ satisfying

$$\mathbf{H}_z^{(2)} = \boldsymbol{\Gamma}\ \mathbf{H}_z^{(1)},$$

and we can rewrite (28) to

$$\underbrace{\begin{bmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Gamma} \end{bmatrix} \mathbf{y}_t^{(1)}}_{:=\tilde{\mathbf{y}}_t = \begin{bmatrix} \tilde{\mathbf{y}}_t^{(1)} \\ \tilde{\mathbf{y}}_t^{(2)} \end{bmatrix}} = \underbrace{\begin{bmatrix} \mathbf{H}_x^{(1)} \\ \mathbf{H}_x^{(2)} - \boldsymbol{\Gamma}\ \mathbf{H}_x^{(1)} \end{bmatrix}}_{:=\tilde{\mathbf{H}}_x} \mathbf{x}_t + \underbrace{\begin{bmatrix} \mathbf{H}_z^{(1)} \\ \mathbf{0} \end{bmatrix}}_{:=\tilde{\mathbf{H}}_z} \mathbf{z}_t. \tag{29}$$

It becomes obvious from (29) that $\tilde{\mathbf{y}}_t^{(2)}$, which is the lower part of the transformed observations vector $\tilde{\mathbf{y}}_t$, only depends on $\mathbf{x}_t$, which was determined in the previous period $t-1$. Hence information on $\tilde{\mathbf{y}}_t^{(2)}$, from the model's perspective, is already available in $t-1$. A possible solution to this problem might be to replace $\tilde{\mathbf{y}}_t^{(2)}$ with $\tilde{\tilde{\mathbf{y}}}_t^{(2)} := \tilde{\mathbf{y}}_{t+1}^{(2)}$. However, in summary, we can state that a singular $\mathbf{H}_z$ matrix in most cases is due to a misspecified observations vector $\mathbf{y}_t$.

Non-stationary state-space models: In cases where the initial state vector $\mathbf{w}_0$ contains non-stationary elements, there is typically no stabilizing solution to the RDE (8a), and it remains unclear if we can meet the preconditions of Assumption 3.1. However, this does not mean that Algorithm 1 is impractical for non-stationary SSMs. To see this, remember that it follows from Proposition 3.1 that for the SSM (2) with $n_y = n_z = \text{rk}(\mathbf{H}_z)$, there is a solution $\mathbf{C}_+ = \mathbf{0}$ satisfying Assumption 3.1. This can be seen as an advantage compared to the CR discussed by Herbst (2015), which strictly requires the transition matrix $\mathbf{F}$ to be stable.

4 Application

In this section, we illustrate the usage of the ASKF and compare it in terms of speed and performance to three competitors. As a benchmark algorithm, we use the regular KF (4), representing one of the most basic versions of the filter. The second competitor is a version of the Chandrasekhar recursion (CR) developed by Morf (1974) and Morf et al. (1974). Compared to the regular KF, this algorithm replaces the RDE (8a) or (8b), respectively, with another set of difference equations. Herbst (2015) points out that this set of "Chandrasekhar-type" difference equations requires fewer arithmetic operations than the regular KF, if the number of states n_w is large compared to the dimension n_y of the observation vector $\mathbf{y}_t$. Therefore, Herbst (2015) suggests using the CR when estimating medium to large-scale DSGE models since these models typically possess a large number of state variables and only a handful of observable variables. The implementation of the CR follows the procedure described by Herbst (2015). As the last competitor, we choose a version of the KF based on the univariate treatment of multivariate observation vectors (UKF) by Koopman and Durbin (2000). To briefly recapitulate the basic idea of this method, suppose that H_i represents the ith row of the matrix $\mathbf{H}$ and that $y_{t,i}$, $u_{t,i}$, and h_i denote the ith element of the vectors $\mathbf{y}_t$, $\mathbf{u}_t$, and $\mathbf{h}$, respectively. Under the assumption that $\mathbf{R}$ is a diagonal matrix, with $\mathbf{R} = \text{diag}\left(R_1^2, \ldots, R_{n_y}^2\right)$, Koopman and Durbin (2000) suggest replacing the multivariate measurement equation (1a) with its univariate equivalent

$$y_{t,i} = h_i + H_i \cdot \mathbf{w}_{t,i} + u_{t,i}, \qquad u_{t,i} \sim \text{N}(0, R_i^2) \qquad \forall i = 1, 2, \ldots, n_y,\ \forall t = 1, 2, \ldots, N, \tag{30a}$$

where $\mathbf{w}_{t,i-1} = \mathbf{w}_t$ for all $i = 1, 2, \ldots, n_y$.[15] Subsequently, the corresponding version of the state equation yields

$$\mathbf{w}_{t,i} = \begin{cases} \mathbf{F} \cdot \mathbf{w}_{t-1,n_y} + \mathbf{v}_t, & i = 1, \\ \mathbf{w}_{t,i-1}, & i = 2, 3, \ldots, n_y, \end{cases}, \qquad \forall t = 1, 2, \ldots, N. \tag{30b}$$

Note that (30) can be interpreted as an univariate SSM with n_w states and $N \cdot n_y$ observations, whose log-likelihood function is obtainable employing the KF.[16] Durbin and Koopman (2012, Chapter 6.4.4) show that compared to the multivariate treatment, this univariate approach can significantly reduce the number of arithmetic operations. This is especially true for models where n_y is large since the UKF avoids the inversion of the $n_y \times n_y$ matrix $\mathbf{U}_t$. Instead, the UKF will compute n_y times the inverse of a scalar. For a textbook treatment of the UKF and its implementation, we refer to Durbin and Koopman (2012, Chapter 6.4)

To compare the four filters, we use two frameworks: First, we analyze the generic SSM by Chib and Ramamurthy (2010) as an example of a classic stationary SSM with measurement error. This simulation model essentially represents the SSM (1) with $n_y = 10$ observation variables and $n_w = 5$ state variables, where 60 of the parameters are estimated while treating the remaining parameters as fixed. We use the same (arbitrary) chosen set of data generating parameters as Chib and Ramamurthy (2010). While this generic SSM has no particular economic interpretation, it is an example of a SSM where the number of observable time series ($n_y = 10$) exceeds the number of unobserved states ($n_w = 5$). Therefore, we will also consider the model transformation suggested by Jungbacker and Koopman (2014), which collapses the initially 10×1 observation vector into a new 5×1 observation vector.[17]

Second, we consider the medium-scale DSGE model introduced by Smets and Wouters (2007) as an example for a SSM without measurement error, where we may use Proposition 3.1 to

[15] Note that the UKF is not restricted to cases where $\mathbf{R}$ is a diagonal matrix (see e.g., Durbin and Koopman, 2012, Chapter 6.4.3).

[16] At this point it should be mentioned that the system matrices of (30) depend on the index i. To deal with this the Kalman recursion (4) must be slightly adjusted.

[17] Briefly summarized, the idea behind this procedure is to find matrices $\mathbf{A}^* \in \mathbb{R}^{n_w \times n_w}$ and $\mathbf{A}^+ \in \mathbb{R}^{n_y - n_w \times n_w}$ to

obtain $\mathbf{C}_+$. Since both the generic SSM by Chib and Ramamurthy (2010) as well as the model by Smets and Wouters (2007) represent stationary models, we use the unconditional initialization strategy to obtain $(\boldsymbol{\mu}_0, \mathbf{C}_0)$.

To obtain the steady-state variance matrix $\mathbf{C}_+$ in the case of the generic SSM by Chib and Ramamurthy (2010), we will use the Schur algorithm described by Bini et al. (2012, Chapter 3) to solve the DARE (9a) corresponding to RDE (8a). We use the same algorithm to solve the discrete Lyapunov equation (7b) for the unconditional variance matrix $\mathbf{C}$ of the state vector $\mathbf{w}_t$.

All computations in this section were coded in MATLAB® 2019a or FORTRAN (using the Intel® IFORT compiler) and executed on a Window 10 64-bit machine with a 3.60 GHz Intel® Core™i7-7700 CPU and 32 GB of RAM. Further, it is worth mentioning that the FORTRAN code makes extensive use of the BLAS and LAPACK routines, such as `dsymm` or `dsryk`, that come with Intel®'s Math Kernel Library to exploit the symmetric nature of variance matrices wherever possible.

To compare the different filters, we consider a Bayesian setup and use the tailored randomized block Metropolis-Hastings (TaRBMH) sampler to generate 11000 draws from the posterior distribution $f_{\boldsymbol{\theta}|\mathbf{Y}_t} \propto f_{\mathbf{Y}_t|\boldsymbol{\theta}} \times f_{\boldsymbol{\theta}}$, where we discard the first 1000 draws as burn-ins. Subsequently, we compare the speed and accuracy of each filter by recomputing each of the 10000 remaining parameter sets using the regular KF as the benchmark.

In a nutshell, we may summarize the TaRBMH sampler by Chib and Ramamurthy (2010) as follows: With each draw from the posterior distribution, we partition the parameter vector $\boldsymbol{\theta}$ into multiple blocks. Thereby, both the number of blocks and the allocation of the parameters into the blocks are random. The parameters of each block are then sequentially updated by a Metropolis-Hastings step, where we draw the proposals from a multivariate student-t density with ν degrees of freedom. To parameterize the proposal density, we follow Chib and Ramamurthy (2010) and detect the conditional posterior mode (with respect to the block-parameters) by means of simulated annealing based on a linear cooling schedule. The proposal density's mean vector and scaling matrix then reflect the conditional posterior mode and the

linearly transform $\mathbf{y}_t$ to

$$\begin{pmatrix}\mathbf{y}_t^*\\ \mathbf{y}_t^+\end{pmatrix} := \begin{pmatrix}\mathbf{A}^*\\ \mathbf{A}^+\end{pmatrix} \mathbf{y}_t,$$

such that we may write the transformed measurement equation as

$$\begin{pmatrix}\mathbf{y}_t^*\\ \mathbf{y}_t^+\end{pmatrix} = \begin{pmatrix}\mathbf{h}^*\\ \mathbf{h}^+\end{pmatrix} + \begin{pmatrix}\mathbf{H}^*\\ \mathbf{0}\end{pmatrix}\mathbf{w}_t + \begin{pmatrix}\mathbf{u}_t^*\\ \mathbf{u}_t^+\end{pmatrix}, \qquad \begin{pmatrix}\mathbf{u}_t^*\\ \mathbf{u}_t^+\end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{R}_t^* & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_t^+\end{pmatrix}\right), \qquad \forall t = 1, 2, \ldots, N.$$

Thus, the collapsed measurement equations yields

$$\mathbf{y}_t^* = \mathbf{h}^* + \mathbf{H}^*\,\mathbf{w}_t + \mathbf{u}_t^*, \qquad \mathbf{u}_t^* \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_t^*), \qquad \forall t = 1, 2, \ldots, N.$$

The vector $\mathbf{y}_t^+ = \mathbf{h}^+ + \mathbf{u}_t^+$ with $\mathbf{u}_t^+ \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_t^+)$, however, is independent of $\mathbf{w}_t$ and $\mathbf{u}_t^*$ and can therefore be treated separately. For a textbook treatment we refer to Durbin and Koopman (2012, Chapter 6.5).

corresponding Hessian matrix.

The tuning parameters of the TaRBMH sampler and the corresponding simulated annealing algorithm (see Table 2) used to specify the proposal density are identical to the setup by Chib and Ramamurthy (2010).

Table 2: TaRBMH and simulated annealing settings

	TaRBMH		
Parameter	Description	GSSM[a]	SW07[b]
p_B	Probability for a new block	0.15	0.15
M	Number of draws	10000	10000
n_0	Number of burn-ins	1000	1000
ν	Degrees of freedom of the proposal density[a]	15	10
	Simulated annealing with linear cooling schedule		
t_0	Initial temperature	5	5
a	Cooling constant	0.4	0.4
K	Number of stages in cooling schedule	8	4
b	Stage expansion factor	8	6
s	Scaling factor for new proposals	0.02	0.02

[a] GSSM: The generic SSM introduced by Chib and Ramamurthy (2010).
[b] SW07: The DSGE model introduced by Smets and Wouters (2007).

4.1 Generic state-space model

We may express the generic SSM by Chib and Ramamurthy (2010) in terms of the SSM (1) by defining the system matrices $\mathbf{h}$, $\mathbf{H}$, $\mathbf{F}$, $\mathbf{Q}$, and $\mathbf{R}$ as

$$\mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_{10} \end{pmatrix} \in \mathbb{R}^{10\times 1}, \qquad \mathbf{H} = \begin{pmatrix} 1 & H_{2,1} & & \cdots & & H_{10,1} \\ & \ddots & \ddots & & & \vdots \\ & & 1 & H_{6,5} & \cdots & H_{10,5} \end{pmatrix}^T \in \mathbb{R}^{10\times 5},$$

$$\mathbf{F} = \operatorname{diag}(F_{1,1}, \dots, F_{5,5}) \in \mathbb{R}^{5\times 5}, \qquad \mathbf{Q} = \mathbf{I} \in \mathbb{R}^{5\times 5}, \quad \mathbf{R} = \operatorname{diag}(e^{\sigma_1^2}, \dots, e^{\sigma_{10}^2}) \in \mathbb{R}^{10\times 10}.$$

Further, the vector

$$\boldsymbol{\theta} = \begin{pmatrix} F_{1,1} & \dots & F_{5,5} & h_1 & \dots & h_{10} & H_{2,1} & \dots & H_{10,5} & \sigma_1^2 & \dots & \sigma_{10}^2 \end{pmatrix}^T$$

collects the 60 uncertain parameters of the model. To estimate the model, we follow Chib and Ramamurthy (2010) and simulate a set of 200 observations using the data generating parameters presented in Table 3. In choosing the prior distributions $f_{\boldsymbol{\theta}}$ of the uncertain parameters displayed in Table 3, we once again follow Chib and Ramamurthy (2010). We report the estimation results for each of the four filters in Table 4. The whole estimation procedure requires about 46 million likelihood evaluations.[18]

Table 5 shows the time needed by each filter to reevaluate the log-densities $f_{\mathbf{Y}_t|\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}$ for all 10000 draws from the posterior distribution. To get an intuition for the numerical accuracy of each filter, Table 5 also provides the l^2-Norm of the deviations between the log-likelihood computed with a particular filter and the log-likelihood evaluated using the regular KF.

For MATLAB® implementation of the full model, we see that, in comparison to the regular KF or the UKF, the ASKF requires less than half of the time. In FORTRAN, the ASKF reduces the computational burden, even more, requiring only about 12 percent and 33 percent of the time compared to the KF and UKF, respectively. Unsurprisingly, the slowest filter for both implementations is the CR. In line with the results of Herbst (2015), we find that the CR becomes inefficient compared to the regular KF when $n_y \geq n_w$.

The lower part of Table 5 displays the results obtained when using the technique described by Jungbacker and Koopman (2014) to collapse the observations vector to the dimension of the state vector. Except for the ASKF, this model transformation significantly reduces the computational burden of all filters. However, the ASKF remains the fastest filter in both the MATLAB® and the FORTRAN implementation.

Overall, the FORTRAN implementation of the generic SSM by Chib and Ramamurthy (2010) seems to be twice as fast as its MATLAB® counterpart. The numerical deviation of the filters compared to the standard KF are similar, with the CR being closest to the KF.

At this point, we have to mention that all results considered so far were under the hypothesis that the convergence process from $\mathbf{C}_0$ to $\mathbf{C}_+$ stretches over the complete observation interval ($N = 200$). However, in practice, there might be some period τ where $\mathbf{C}_\tau$ has converged sufficiently close to $\mathbf{C}_+$ so that we can switch from the filter at hand to the SKF described in (11). Thus, if the convergence process lasts only a few periods, it could be that the additional effort to solve the RDE (8a) for $\mathbf{C}_+$ outweighs the efficiency gains from the more efficient recursive part of the ASKF. To get an intuition of how many periods are necessary so that the ASKF outperforms the other filters in terms of speed, depending on the number of observations N, Figure 1 displays the computation time of each of the filters relative to the computation time of the KF.

As expected, Figure 1 shows that for a low number of observations ($N < 10$), the ASKF is outperformed by the other filters but becomes faster as N rises. For the MATLAB® implementation of the full model, the ASKF becomes the fastest option to compute the log-likelihood in cases where the convergence process takes more than 25 periods, while in FORTRAN, it takes the ASKF about 50 periods to outperform the UKF. When using the collapsed model, where the efficiency gains from the ASKF are smaller, in both MATLAB® and the FORTRAN implementation,

[18] A large part of these likelihood evaluations stem from the repeated numerical evaluation of the conditional Hessian matrix, which is required to specify the proposal density of a random block.

it takes about 75 periods for the ASKF to become the fastest option.

Since, depending on the model's parameter values, the convergence speed of the matrix sequence $\{\mathbf{C}_t\}_{t=1}^{N}$ may vary, in practice, it is often impossible to determine ex-ante in which period a switch to SKF is possible. Thus, using the ASKF to evaluate the log-likelihood is probably not a bad choice, especially considering that in cases where we may switch early to the SKF, the choice of the filter might become secondary for the overall time needed to evaluate the log-likelihood.

Table 3: Data generating parameters and prior density – Generic state-space model

Parameter	Data generating parameters					Prior[a]				
$h_1, \ldots, h_5$	0.20	1.40	1.80	0.10	0.90	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)
$h_6, \ldots, h_{10}$	1.00	2.00	0.10	2.20	1.50	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)	0.50 (5.00)
$H_{2,1}$	0.50					0.00 (5.00)				
$H_{3,1}, H_{3,2}$	0.60	0.00				0.00 (5.00)	0.00 (5.00)			
$H_{4,1}, \ldots, H_{4,3}$	0.00	0.20	−0.10			0.00 (5.00)	0.00 (5.00)	0.00 (5.00)		
$H_{5,1}, \ldots, H_{5,4}$	−0.20	0.00	−0.70	0.00		0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	
$H_{6,1}, \ldots, H_{6,5}$	0.00	0.00	−0.40	−0.50	0.00	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)
$H_{7,1}, \ldots, H_{7,5}$	0.30	0.20	0.00	0.00	−0.30	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)
$H_{8,1}, \ldots, H_{8,5}$	−0.50	0.00	0.00	0.60	0.00	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)
$H_{9,1}, \ldots, H_{9,5}$	0.00	−0.50	0.30	−0.10	0.00	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)
$H_{10,1}, \ldots, H_{10,5}$	0.00	0.00	0.20	0.00	−0.40	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)	0.00 (5.00)
$F_{1,1}, \ldots, F_{5,5}$	0.80	0.20	0.75	0.60	0.10					
$\sigma_1^2, \ldots, \sigma_5^2$	log(1.00)	log(0.30)	log(1.00)	log(0.20)	log(0.60)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)
$\sigma_6^2, \ldots, \sigma_{10}^2$	log(0.50)	log(1.00)	log(1.00)	log(0.75)	log(0.60)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)	−1.00 (1.00)

[a] All parameters are normally distributed. The first parameter denotes the prior mean, while the second parameter (in parentheses) denotes the prior variance.

4.2 Smets and Wouters model

The model introduced by Smets and Wouters (2003, 2007) is at the core of most medium- to large-scale DSGE models used to analyze monetary policy. To put it in the words of Herbst and Schorfheide (2016, pp. 12): „By now, the SW model has become one of the workhorse models in the DSGE model literature and in central banks around the world."

Among other features, the model includes sticky prices and wages, investment adjustment costs, habit formation, and variable capital utilization. In what follows, we use the Smets and

Table 4: Estimation results – Generic state-space model

$\boldsymbol{\theta}$	Mean				5 percent quantile				95 percent quantile			
	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
$F_{1,1}$	0.74	0.74	0.74	0.74	0.64	0.64	0.64	0.64	0.84	0.83	0.84	0.84
$F_{2,2}$	0.36	0.36	0.36	0.36	0.19	0.19	0.19	0.19	0.53	0.53	0.53	0.53
$F_{3,3}$	0.70	0.70	0.70	0.70	0.58	0.58	0.58	0.58	0.81	0.81	0.81	0.81
$F_{4,4}$	0.41	0.41	0.42	0.42	0.27	0.27	0.28	0.28	0.55	0.56	0.56	0.56
$F_{5,5}$	0.06	0.06	0.06	0.06	−0.15	−0.15	−0.15	−0.15	0.28	0.27	0.27	0.28
h_1	0.15	0.16	0.14	0.15	−0.40	−0.38	−0.43	−0.41	0.69	0.68	0.70	0.71
h_2	1.20	1.21	1.20	1.20	0.88	0.88	0.86	0.87	1.51	1.52	1.53	1.52
h_3	1.65	1.68	1.64	1.66	1.11	1.04	1.02	1.06	2.25	2.30	2.26	2.27
h_4	0.22	0.22	0.22	0.21	−0.06	−0.07	−0.06	−0.06	0.49	0.50	0.50	0.48
h_5	0.89	0.88	0.90	0.89	0.52	0.48	0.51	0.50	1.27	1.30	1.29	1.26
h_6	0.94	0.94	0.95	0.95	0.72	0.71	0.71	0.72	1.17	1.17	1.18	1.18
h_7	1.90	1.90	1.90	1.90	1.67	1.66	1.67	1.67	2.12	2.12	2.12	2.12
h_8	0.33	0.32	0.33	0.32	0.07	0.07	0.07	0.06	0.57	0.59	0.59	0.57
h_9	2.10	2.10	2.10	2.10	1.89	1.88	1.88	1.89	2.32	2.32	2.32	2.32
h_{10}	1.50	1.50	1.50	1.50	1.36	1.35	1.35	1.36	1.64	1.65	1.65	1.65
$H_{2,1}$	0.44	0.44	0.44	0.44	0.28	0.29	0.29	0.28	0.60	0.60	0.60	0.60
$H_{3,1}$	0.69	0.69	0.70	0.69	0.46	0.46	0.47	0.46	0.92	0.93	0.93	0.92
$H_{4,1}$	0.14	0.14	0.14	0.13	−0.02	−0.03	−0.05	−0.04	0.31	0.31	0.31	0.30
$H_{5,1}$	−0.30	−0.30	−0.31	−0.30	−0.48	−0.50	−0.50	−0.49	−0.11	−0.11	−0.12	−0.12
$H_{6,1}$	−0.08	−0.08	−0.08	−0.08	−0.22	−0.21	−0.22	−0.21	0.05	0.06	0.06	0.06
$H_{7,1}$	0.31	0.31	0.31	0.31	0.20	0.20	0.20	0.20	0.42	0.42	0.43	0.42
$H_{8,1}$	−0.27	−0.28	−0.28	−0.28	−0.42	−0.43	−0.43	−0.43	−0.13	−0.13	−0.12	−0.13
$H_{9,1}$	0.07	0.07	0.07	0.08	−0.06	−0.05	−0.05	−0.05	0.20	0.20	0.20	0.20
$H_{10,1}$	0.03	0.03	0.03	0.03	−0.06	−0.06	−0.06	−0.06	0.13	0.13	0.13	0.13
$H_{3,2}$	−0.06	−0.06	−0.06	−0.05	−0.32	−0.33	−0.33	−0.33	0.20	0.22	0.21	0.22
$H_{4,2}$	0.21	0.21	0.21	0.22	−0.00	−0.01	−0.01	0.00	0.44	0.44	0.45	0.46
$H_{5,2}$	−0.00	−0.00	−0.00	−0.01	−0.26	−0.28	−0.27	−0.28	0.25	0.26	0.26	0.26
$H_{6,2}$	0.18	0.18	0.18	0.17	0.01	0.00	−0.01	−0.00	0.36	0.35	0.36	0.35
$H_{7,2}$	0.14	0.14	0.14	0.14	−0.04	−0.04	−0.04	−0.04	0.31	0.32	0.31	0.31
$H_{8,2}$	−0.05	−0.05	−0.05	−0.05	−0.26	−0.26	−0.26	−0.25	0.16	0.15	0.16	0.16
$H_{9,2}$	−0.57	−0.57	−0.57	−0.57	−0.75	−0.75	−0.75	−0.75	−0.39	−0.39	−0.39	−0.39
$H_{10,2}$	−0.11	−0.10	−0.11	−0.10	−0.25	−0.25	−0.25	−0.25	0.03	0.04	0.04	0.04
$H_{4,3}$	−0.21	−0.20	−0.20	−0.19	−0.39	−0.39	−0.39	−0.39	−0.01	−0.01	−0.02	−0.00
$H_{5,3}$	−0.63	−0.63	−0.63	−0.63	−0.80	−0.80	−0.80	−0.81	−0.46	−0.46	−0.46	−0.46
$H_{6,3}$	−0.34	−0.34	−0.34	−0.35	−0.48	−0.47	−0.47	−0.49	−0.22	−0.21	−0.21	−0.22
$H_{7,3}$	−0.11	−0.11	−0.11	−0.11	−0.23	−0.24	−0.24	−0.23	0.01	0.01	0.01	0.01
$H_{8,3}$	−0.14	−0.14	−0.14	−0.14	−0.30	−0.30	−0.30	−0.30	0.03	0.03	0.02	0.03
$H_{9,3}$	0.28	0.28	0.28	0.28	0.17	0.17	0.17	0.17	0.39	0.39	0.39	0.39
$H_{10,3}$	0.17	0.17	0.17	0.17	0.08	0.07	0.07	0.07	0.27	0.27	0.26	0.27
$H_{5,4}$	0.09	0.08	0.09	0.09	−0.12	−0.11	−0.11	−0.11	0.28	0.28	0.28	0.29
$H_{6,4}$	−0.58	−0.58	−0.58	−0.58	−0.70	−0.70	−0.70	−0.70	−0.47	−0.46	−0.47	−0.47
$H_{7,4}$	−0.00	0.00	0.00	0.00	−0.14	−0.13	−0.14	−0.13	0.14	0.14	0.14	0.14
$H_{8,4}$	0.61	0.61	0.61	0.61	0.47	0.47	0.47	0.47	0.76	0.76	0.76	0.76
$H_{9,4}$	−0.11	−0.11	−0.11	−0.11	−0.23	−0.23	−0.23	−0.23	0.01	0.01	0.01	0.01
$H_{10,4}$	−0.01	−0.01	−0.01	−0.01	−0.13	−0.12	−0.12	−0.12	0.10	0.10	0.10	0.10
$H_{6,5}$	0.19	0.19	0.19	0.19	0.05	0.05	0.05	0.05	0.33	0.33	0.33	0.33
$H_{7,5}$	−0.41	−0.42	−0.42	−0.42	−0.61	−0.61	−0.61	−0.61	−0.23	−0.23	−0.23	−0.23
$H_{8,5}$	−0.03	−0.02	−0.03	−0.02	−0.22	−0.21	−0.22	−0.21	0.16	0.17	0.16	0.17
$H_{9,5}$	−0.18	−0.18	−0.18	−0.18	−0.34	−0.33	−0.34	−0.34	−0.02	−0.02	−0.02	−0.02
$H_{10,5}$	−0.49	−0.49	−0.49	−0.49	−0.64	−0.65	−0.65	−0.65	−0.34	−0.34	−0.34	−0.34
σ_1^2	0.08	0.09	0.09	0.08	−0.23	−0.23	−0.23	−0.24	0.39	0.40	0.40	0.38
σ_2^2	−0.69	−0.70	−0.70	−0.68	−1.31	−1.34	−1.30	−1.28	−0.21	−0.22	−0.22	−0.21
σ_3^2	−0.34	−0.35	−0.35	−0.34	−0.81	−0.83	−0.82	−0.81	0.07	0.05	0.05	0.07
σ_4^2	−1.44	−1.45	−1.43	−1.44	−2.30	−2.31	−2.24	−2.29	−0.81	−0.81	−0.80	−0.80
σ_5^2	−0.15	−0.14	−0.15	−0.15	−0.58	−0.57	−0.56	−0.56	0.20	0.21	0.21	0.21
σ_6^2	−0.97	−0.96	−0.97	−0.97	−1.29	−1.28	−1.29	−1.29	−0.68	−0.68	−0.69	−0.68
σ_7^2	−0.10	−0.10	−0.10	−0.10	−0.33	−0.34	−0.33	−0.34	0.12	0.12	0.11	0.12
σ_8^2	0.07	0.07	0.07	0.07	−0.13	−0.13	−0.13	−0.13	0.27	0.28	0.28	0.27
σ_9^2	−0.48	−0.48	−0.48	−0.49	−0.75	−0.75	−0.75	−0.76	−0.23	−0.23	−0.24	−0.24
σ_{10}^2	−0.72	−0.73	−0.72	−0.73	−1.00	−1.02	−1.01	−0.99	−0.47	−0.48	−0.48	−0.49

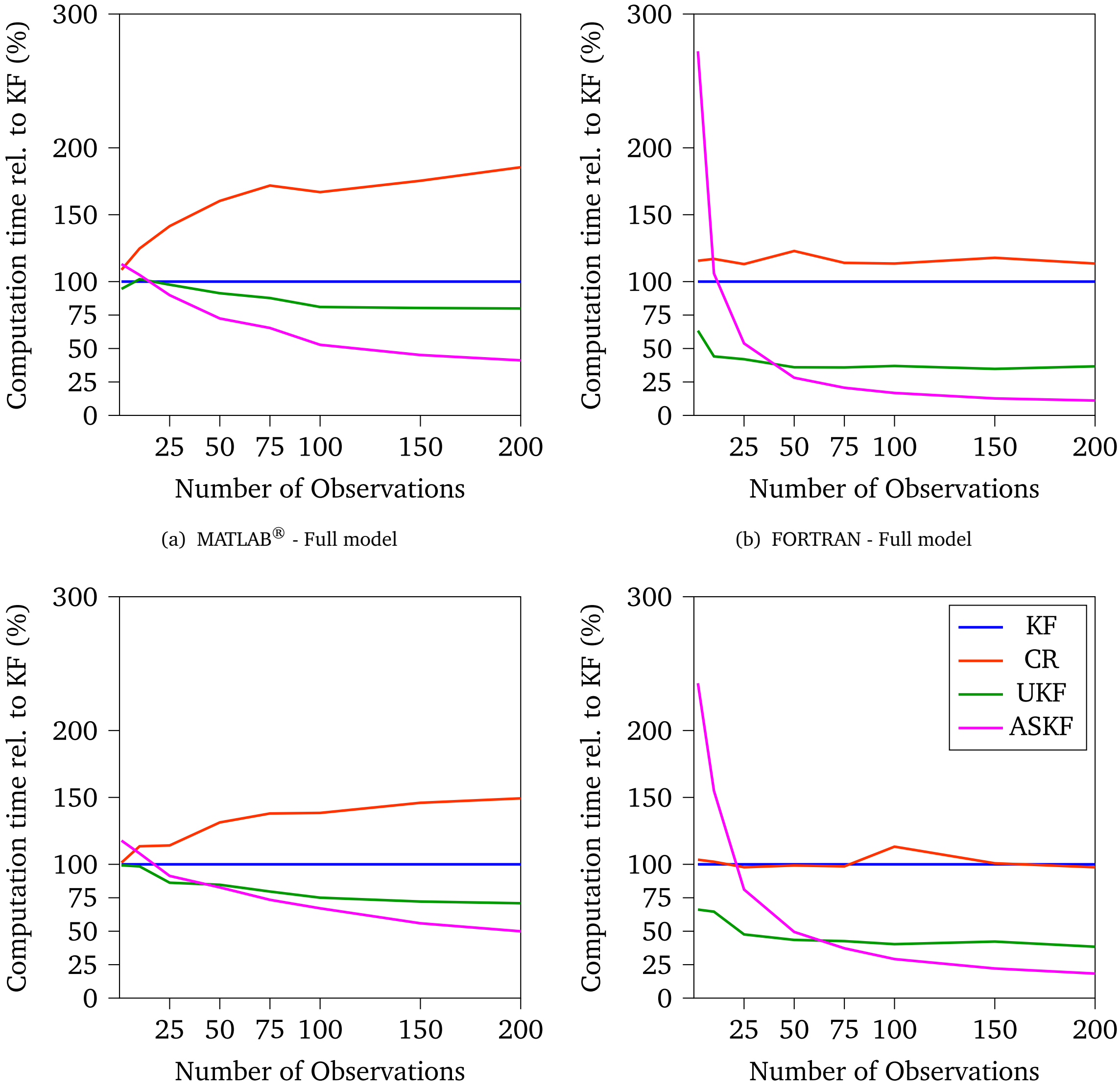
Wouters (2007) model, which is a slightly adjusted version of the original model described by Smets and Wouters (2003). The model consists of 62 equations in 35 endogenous variables, 20 predetermined, and 7 endogenous state variables. We give a complete description of the

Table 5: Speed comparison – Generic state-space model

Full	MATLAB®				FORTRAN			
$n_y = 10,\ n_w = 5$	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
Elapsed time:	$43s$	$82s$	$35s$	$17s$	$17s$	$19s$	$6s$	$2s$
l^2-Norm:	-	$2.1e^{-09}$	$2.2e^{-09}$	$2.1e^{-09}$	-	$0.5e^{-10}$	$0.6e^{-09}$	$0.2e^{-09}$

Collapsed	MATLAB®				FORTRAN			
$n_y = 5,\ n_w = 5$	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
Elapsed time:	$32s$	$46s$	$22s$	$16s$	$9s$	$9s$	$4s$	$2s$
l^2-Norm:	-	$1.1e^{-09}$	$1.2e^{-09}$	$1.1e^{-09}$	-	$0.3e^{-10}$	$0.2e^{-09}$	$0.1e^{-09}$

Figure 1: Speed comparison – Generic state-space model



(a) MATLAB® - Full model

(b) FORTRAN - Full model

(c) MATLAB® - Collapsed model

(d) FORTRAN - Collapsed model

model's implementation in Appendix E. To fit the model to the data, we follow Smets and Wouters (2007) and use quarterly time series of the log difference of real GDP, the log difference of real consumption, the log difference of real investment, and the log difference of real wages, the log of hours worked, the log difference of GDP deflator, and the federal funds rate for the U.S. from 1966 : 1 to 2004 : 4.

To obtain the model's linear policy function, we use the general Schur decomposition in the manner of Klein (2000). We then transform the solved model into a SSM without a measurement error. We consider two different state-space representations of the model: (i) A reduced SSM with $n_y = 7$ and $n_w = 27$, including only the predetermined and endogenous state variables in the state vector, and (ii) the full SSM with $n_y = 7$ and $n_w = 62$, treating all of the model's variables as state variables.[19] Further, as shown in Appendix E, the log-linearized model satisfies the preconditions of Proposition 3.1, and thus we may obtain the solution of the RDE (8a) as $\mathbf{C}_+ = \mathbf{0}$.

To estimate the model, we use the same prior densities as Smets and Wouters (2007). In Table 6, we report these prior densities together with the estimation results of the four different filters. All results of Table 6 refer to the reduced SSM and were computed in MATLAB®. Again, the posterior statistics obtained by the different filters are similar. Each filter requires about 15 million likelihood evaluations to generate the 10000 draws from the posterior distribution. The filters also perform similarly in terms of the percentages of failed tries, where the particular algorithm could not evaluate the objective function.

The most significant differences between the filters occur in the total time required to estimate the model. The ASKF is about one-third faster than the standard KF and reduces the overall estimation time by more than 11 hours. Moreover, compared to the UKF, which represents the second fastest option in this setup, the ASKF is still 6 hours ahead. This is despite the fact that the evaluation of the log-likelihood now also includes the computation of the policy function so that the actual filtering process only accounts for a part of the total evaluation time. Surprisingly, the CR performs even slower than the KF, which leads to the conclusion that in the reduced version of the Smets and Wouters model, n_w is too small compared to n_y for the CR to be efficient.

Table 7 shows the results of the reevaluation of the 10000 draws from the posterior distribution. In addition to the time required for the actual filtering process, we also report the times required to compute the policy function, the unconditional initialization, and the prior density. Considering the reduced model, we again see that the ASKF significantly reduces the computational time required in both MATLAB® and FORTRAN. In MATLAB®, the ASKF reduces the actual filtering time compared to the KF, the CR, and the UKF by about 73, 79, and 60 percent,

[19]Note that using the full SSM will result in a singular unconditional variance matrix $\mathbf{C}$ of the state vector $\mathbf{w}_t$, since in this case $\mathbf{w}_t$ contains redundant states.

respectively. In FORTRAN, the computation time is reduced by 82 and 69 percent, respectively, compared to the KF and the UKF. Compared to its MATLAB® implementation, the CR performs better in FORTRAN. Nevertheless, compared to the ASKF, the computational burden is almost four times higher.

The lower part of Table 7 displays the results of the speed comparison in case we include all 62 variables of the Smets and Wouters model as states in the SSM. The main difference in this setup is the performance of the CR, which is now about twice as fast as the KF and UKF, respectively, in both MATLAB® and FORTRAN. The CR also gets closest to the ASKF in terms of speed, but it is still three times faster in MATLAB® and almost twice as fast in FORTRAN.

In all four implementations considered in Table 7, the ASKF reduces the portion of the actual filtering process on the total computing time to less than one-fifth. Furthermore, it is worth mentioning that in the case of the Smets and Wouters model, the performance of the ASKF is also less dependent on the convergence speed of the matrix sequence $\{\mathbf{C}_t\}_{t=1}^{N}$, since due to Proposition 3.1, we do not have to determine $\mathbf{C}_+$ numerically.

Table 6: Prior and estimation results – Smets and Wouters model

$\boldsymbol{\theta}$	Prior			Posterior											
				Mean				5 percent quantile				95 percent quantile			
	Density[a]	Mean	Std. Dev.	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
φ	Normal	4.000	1.500	5.60	5.47	5.54	5.48	3.63	3.40	3.60	3.40	8.04	7.91	7.93	7.98
σ_c	Normal	1.500	0.375	1.33	1.32	1.31	1.33	1.11	1.09	1.10	1.11	1.59	1.57	1.56	1.59
h	Beta	0.700	0.100	0.72	0.72	0.72	0.71	0.63	0.63	0.64	0.62	0.79	0.80	0.80	0.80
ξ_w	Beta	0.500	0.100	0.70	0.69	0.70	0.69	0.57	0.57	0.57	0.57	0.81	0.80	0.82	0.80
σ_l	Normal	2.000	0.750	1.83	1.84	1.85	1.79	0.82	0.79	0.84	0.77	3.13	3.15	3.08	3.05
ξ_p	Beta	0.500	0.100	0.64	0.63	0.64	0.64	0.54	0.54	0.55	0.54	0.75	0.73	0.75	0.74
ι_w	Beta	0.500	0.150	0.59	0.58	0.58	0.58	0.34	0.34	0.36	0.33	0.80	0.82	0.80	0.80
ι_p	Beta	0.500	0.150	0.24	0.25	0.24	0.25	0.10	0.11	0.10	0.11	0.41	0.44	0.42	0.42
ψ	Beta	0.500	0.150	0.56	0.57	0.56	0.56	0.34	0.35	0.36	0.35	0.77	0.77	0.75	0.76
Φ	Normal	1.250	0.125	1.60	1.60	1.60	1.60	1.46	1.46	1.46	1.46	1.75	1.74	1.74	1.75
r_π	Normal	1.500	0.250	2.03	2.05	2.03	2.06	1.71	1.75	1.72	1.73	2.39	2.38	2.37	2.40
ρ	Beta	0.750	0.100	0.80	0.80	0.80	0.80	0.75	0.75	0.75	0.75	0.84	0.84	0.84	0.85
r_y	Normal	0.125	0.050	0.08	0.08	0.08	0.09	0.05	0.05	0.05	0.05	0.13	0.13	0.13	0.13
$r_{\Delta y}$	Normal	0.125	0.050	0.22	0.22	0.22	0.22	0.17	0.17	0.17	0.17	0.27	0.27	0.27	0.27
$\bar{\pi}$	Gamma	0.625	0.100	0.71	0.70	0.71	0.71	0.51	0.52	0.52	0.52	0.91	0.91	0.91	0.91
$\tilde{\beta}$	Gamma	0.250	0.100	0.16	0.16	0.17	0.16	0.07	0.07	0.08	0.07	0.28	0.27	0.27	0.27
$\bar{l}$	Normal	0.000	2.000	0.73	0.84	0.84	0.74	−1.87	−1.60	−1.71	−1.77	3.30	3.29	3.22	3.31
$\bar{\gamma}$	Normal	0.400	0.100	0.42	0.42	0.42	0.42	0.39	0.38	0.39	0.39	0.45	0.45	0.45	0.45
α	Normal	0.300	0.050	0.19	0.19	0.19	0.19	0.16	0.16	0.16	0.16	0.22	0.22	0.22	0.22
ρ_a	Beta	0.500	0.200	0.96	0.96	0.96	0.96	0.93	0.94	0.93	0.94	0.98	0.98	0.98	0.98
ρ_b	Beta	0.500	0.200	0.24	0.25	0.24	0.26	0.08	0.09	0.08	0.08	0.48	0.45	0.46	0.51
ρ_g	Beta	0.500	0.200	0.98	0.98	0.98	0.98	0.96	0.96	0.96	0.96	0.99	0.99	0.99	0.99
ρ_i	Beta	0.500	0.200	0.71	0.72	0.72	0.71	0.60	0.60	0.61	0.60	0.82	0.83	0.83	0.82
ρ_r	Beta	0.500	0.200	0.16	0.16	0.16	0.16	0.06	0.06	0.06	0.05	0.29	0.29	0.28	0.28
ρ_p	Beta	0.500	0.200	0.90	0.90	0.89	0.90	0.79	0.80	0.78	0.79	0.97	0.97	0.97	0.98
ρ_w	Beta	0.500	0.200	0.97	0.97	0.97	0.97	0.94	0.95	0.94	0.94	0.99	0.99	0.99	0.99
μ_p	Beta	0.500	0.200	0.69	0.70	0.68	0.70	0.45	0.46	0.45	0.49	0.84	0.87	0.85	0.86
μ_w	Beta	0.500	0.200	0.85	0.85	0.84	0.84	0.72	0.71	0.70	0.70	0.94	0.94	0.93	0.94
ρ_{ga}	Beta	0.500	0.250	0.52	0.52	0.51	0.52	0.34	0.35	0.34	0.34	0.69	0.69	0.69	0.69
σ_a	Inv. Gamma	0.100	2.000	0.46	0.46	0.46	0.46	0.41	0.41	0.41	0.41	0.51	0.51	0.51	0.51
σ_b	Inv. Gamma	0.100	2.000	0.24	0.24	0.24	0.23	0.18	0.19	0.19	0.18	0.28	0.28	0.29	0.28
σ_g	Inv. Gamma	0.100	2.000	0.53	0.53	0.53	0.53	0.48	0.48	0.48	0.48	0.58	0.59	0.59	0.59
σ_i	Inv. Gamma	0.100	2.000	0.45	0.45	0.45	0.45	0.37	0.37	0.36	0.37	0.54	0.55	0.54	0.56
σ_r	Inv. Gamma	0.100	2.000	0.25	0.25	0.25	0.25	0.22	0.22	0.22	0.22	0.28	0.28	0.28	0.28
σ_p	Inv. Gamma	0.100	2.000	0.14	0.14	0.14	0.14	0.11	0.11	0.11	0.11	0.17	0.17	0.17	0.17
σ_w	Inv. Gamma	0.100	2.000	0.25	0.25	0.25	0.25	0.21	0.21	0.21	0.21	0.29	0.29	0.29	0.29

	KF	CR	UKF	ASKF
Overall - Estimation time:	$35h\ 05m\ 23s$	$38h\ 43m\ 02s$	$30h\ 11m\ 23s$	$23h\ 46m\ 24s$
Value of Ojc.Fct at Posterior Mode:[b]	−858.14	−858.26	−857.97	−856.63
Number of likelihood evaluations:	15174679	15173491	15193695	15190701
Percentage of failed evaluations:	0.49	0.50	0.47	0.49
Acceptance rate (in %):	49.86	50.12	49.77	49.85

[a] Inv. Gamma denotes the Inverse Gamma type-1 distribution.
[b] Refers to the highest value of the object function for all 10000 draws.

5 Conclusion

The objective of this paper was to propose the ASKF as an efficient algorithm to evaluate the likelihood of linear and time-invariant SSMs. The results concerning the performance of the ASKF are promising. It performs well regardless of whether the number of observable time series n_y outweighs the number of states n_z or vice versa. The basis for its efficiency is the – compared to the regular KF – faster recursive part of the ASKF, reducing the cost per additional observation. The ultimate performance of the ASKF is mainly determined by two factors: The length of the filtering period and the time needed to determine the equilibrium variance matrix of the model's states, where the former is determined by the number of observations N of the available data set and the required periods τ until it might be possible to switch to the SKF. The larger the filtering period, the less the additional computational effort to solve RDE (8a) for $\mathbf{C}_+$ will weigh compared to the total filtering time. Furthermore, as we show in Proposition 3.1, for many DSGE models, such as the model introduced by Smets and Wouters (2007), it is not even necessary to solve RDE (8a) numerically, since for SSMs of the form (2) with $n_y = n_z = \text{rk}(\mathbf{H}_z)$ an analytic solution for $\mathbf{C}_+$ is available. This feature makes the ASKF for these kinds of model even more attractive.

Table 7: Speed comparison – Smets and Wouters model

Reduced	MATLAB®				FORTRAN			
$n_y = 7,\ n_w = 27$	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
Filtering time:	$37s$	$47s$	$25s$	$10s$	$22s$	$15s$	$13s$	$4s$
l^2-Norm:	-	$3.0e^{-08}$	$3.0e^{-09}$	$1.2e^{-10}$	-	$0.8e^{-07}$	$0.1e^{-08}$	$0.2e^{-09}$
Policy function	$17s$				$33s$			
Initialization	$18s$				$18s$			
Prior density	$13s$				$0s$			

Full	MATLAB®				FORTRAN			
$n_y = 7,\ n_w = 62$	KF	CR	UKF	ASKF	KF	CR	UKF	ASKF
Filtering time:	$101s$	$57s$	$92s$	$18s$	$79s$	$28s$	$65s$	$15s$
l^2-Norm:	-	$1.8e^{-08}$	$2.1e^{-09}$	$2.2e^{-09}$	-	$0.9e^{-08}$	$0.9e^{-09}$	$0.4e^{-09}$
Policy function	$18s$				$33s$			
Initialization	$74s$				$71s$			
Prior density	$13s$				$0s$			

REFERENCES

AN, S. AND F. SCHORFHEIDE (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172.

ANDERSON, B. AND J. MOORE (1979): *Optimal Filtering*, Information and System Sciences Series, Englewood Cliffs, NJ: Prentice-Hall.

ANDREASEN, M. M. (2010): "How to Maximize the Likelihood Function for a DSGE Model," *Comput. Econ.*, 35, 127–154.

BINI, D. A., B. IANNAZZO, AND B. MEINI (2012): *Numerical Solution of Algebraic Riccati Equations*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

CAINES, P. E. AND D. Q. MAYNE (1970): "On the Discrete Time Matrix Riccati Equation of Optimal Control," *International Journal of Control*, 12, 785–794.

CHAN, S., G. GOODWIN, AND K. SIN (1984): "Convergence Properties of the Riccati Difference Equation in Optimal Filtering of Nonstabilizable Systems," *IEEE Transactions on Automatic Control*, 29, 110–118.

CHARI, V. V., P. J. KEHOE, AND E. R. MCGRATTAN (2007): "Business Cycle Accounting," *Econometrica*, 75, 781–836.

CHIANG, C.-Y., H.-Y. FAN, AND W.-W. LIN (2010): "A Structured Doubling Algorithm for Discrete-Time Algebraic Riccati Equations with Singular Control Weighting Matrices," *Taiwanese Journal of Mathematics*, 14, 933–954.

CHIB, S. AND S. RAMAMURTHY (2010): "Tailored Randomized Block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155, 19–38.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

DE JONG, P. (1988): "The Likelihood for a State Space Model," *Biometrika*, 75, 165–169.

——— (1991): "The Diffuse Kalman Filter," *The Annals of Statistics*, 19, 1073–1083.

DE NICOLAO, G. AND M. GEVERS (1992): "Difference and Differential Riccati Equations: A Note on the Convergence to the Strong Solution," *IEEE Transactions on Automatic Control*, 37, 1055–1057.

DE SOUZA, C., M. GEVERS, AND G. GOODWIN (1986): "Riccati Equations in Optimal Filtering of Nonstabilizable Systems Having Singular State Transition Matrices," *IEEE Transactions on Automatic Control*, 31, 831–838.

DEJONG, D. N. AND C. DAVE (2011): *Structural Macroeconometrics: Second Edition*, Princeton, NJ: Princeton University Press.

DRYGALLA, A., O. HOLTEMÖLLER, AND K. KIESEL (2020): "The Effects of Fiscal Policy in an Estimated DSGE Model – The Case of the German Stimulus Packages during the Great Recession," *Macroeconomic Dynamics*, 24, 1315–1345.

DURBIN, J. AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Methods: Second Edition*, Oxford Statistical Science Series, Oxford, UK: Oxford University Press.

GADATSCH, N., K. HAUZENBERGER, AND N. STÄHLER (2016): "Fiscal Policy during the Crisis: A Look on Germany and the Euro Area with GEAR," *Economic Modelling*, 52, 997–1016.

GU, G. (2012): *Discrete-Time Linear Systems: Theory and Design with Applications*, New York, NY: Springer.

HAMILTON, J. D. (1994): *Time Series Analysis*, Princeton, NJ: Princeton University Press.

HANSEN, G. D. (1985): "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309–327.

HANSEN, L. P. AND T. J. SARGENT (2013): *Recursive Models of Dynamic Linear Economies*, Princeton, NJ: Princeton University Press.

HARVEY, A. C. (1990a): *The Econometric Analysis of Time Series, 2nd Edition*, MIT Press Books, Cambridge, MA: The MIT Press.

——— (1990b): *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, UK: Cambridge University Press.

HERBST, E. (2015): "Using the "Chandrasekhar Recursions" for Likelihood Evaluation of DSGE Models," *Comput. Econ.*, 45, 693–705.

HERBST, E. AND F. SCHORFHEIDE (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 29, 1073–1098.

HERBST, E. P. AND F. SCHORFHEIDE (2016): *Bayesian Estimation of DSGE Models*, Princeton, NJ: Princeton University Press.

IRELAND, P. N. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

JUNGBACKER, B. AND S. J. KOOPMAN (2014): "Likelihood-Based Dynamic Factor Analysis for Measurement and Forecasting," *The Econometrics Journal*, 18, C1–C21.

KING, R. G., C. I. PLOSSER, AND S. T. REBELO (1988): "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195–232.

KLEIN, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 24, 1405–1423.

KOOPMAN, S. J. AND J. DURBIN (2000): "Fast Filtering and Smoothing for Multivariate State Space Models," *Journal of Time Series Analysis*, 21, 281–296.

LEEPER, E. M., M. PLANTE, AND N. TRAUM (2010): "Dynamics of Fiscal Financing in the United States," *Journal of Econometrics*, 156, 304–321.

LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*, Berlin / Heidelberg, Germany: Springer.

MORF, M. (1974): "Fast Algorithms for Multivariable Systems." Ph.D. thesis, Stanford University, Stanford, CA.

MORF, M., G. SIDHU, AND T. KAILATH (1974): "Some New Algorithms for Recursive Estimation in Constant, Linear, Discrete-Time Systems," *IEEE Transactions on Automatic Control*, 19, 315–323.

ROSENBERG, B. (1973): "Random Coefficients Models: The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression," in *Annals of Economic and Social Measurement, Volume 2, number 4*, ed. by S. V. Berg, NBER, 399–428.

SEARLE, S. R. AND A. I. KHURI (2017): *Matrix Algebra Useful for Statistics*, Wiley Series in Probability and Statistics, Hoboken, NJ: Wiley.

SMETS, F. AND R. WOUTERS (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

——— (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–606.

STOCK, J. AND M. WATSON (2016): "Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, 415–525.

ŠUSTEK, R. (2011): "Monetary Business Cycle Accounting," *Review of Economic Dynamics*, 14, 592–612.

APPENDIX

An Augmented Steady-State Kalman Filter to Evaluate the Likelihood of Linear and Time-Invariant State-Space Models

The Appendix of this paper is structured as follows: The first section contains the formal derivation of the standard Kalman filter (KF) and, in particular, of the difference equations that determine the sequence $\{\mathbf{C}_t\}_{t=1}^N$ of the states' conditional variance matrix. In Section B, we establish the formal foundation to address the question under which conditions the sequence $\{\mathbf{C}_t\}_{t=1}^N$ converges to a long-run equilibrium. Given this long-term equilibrium, we derive the set of equations determining the steady-state Kalman filter (SKF) in Section C. In Section D, we provide the formal derivation of the augmented Kalman filter (AKF), which, together with the SKF, builds the basis of the augmented steady-state Kalman filter (ASKF) proposed in this paper. The last section of the Appendix outlines the implementation of the dynamic stochastic general equilibrium (DSGE) model by Smets and Wouters (2007) that we employ as an application in this paper.

A Derivation of the Kalman filter

This appendix contains the formal derivation of the Kalman recursion (4) with respect to the state-space model (SSM) (1), where for the most part, we will follow the textbook treatments by Durbin and Koopman (2012, Chapter 4) and Harvey (1990b, Chapter 3). For convenience, let us restate the linear, time-invariant, and Gaussian SSM (1) introduced in Section 2:

$$\mathbf{y}_t = \mathbf{h} + \mathbf{H} \cdot \mathbf{w}_t + \mathbf{u}_t, \qquad \mathbf{u}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{R}), \qquad \forall t = 1, 2, \ldots, N$$

$$\mathbf{w}_t = \mathbf{F} \cdot \mathbf{w}_{t-1} + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}), \qquad \mathbf{w}_0 \sim \mathrm{N}(\boldsymbol{\mu}_0, \mathbf{C}_0), \qquad \forall t = 1, 2, \ldots, N$$

with

$$\mathrm{E}[\mathbf{u}_i \mathbf{u}_j^T] = \begin{cases} \mathbf{R}, & i = j, \\ \mathbf{0}, & i \neq j. \end{cases}, \quad \mathrm{E}[\mathbf{v}_i \mathbf{v}_j^T] = \begin{cases} \mathbf{Q}, & i = j, \\ \mathbf{0}, & i \neq j. \end{cases}, \quad \mathrm{E}[\mathbf{u}_i \mathbf{v}_j^T] = \mathbf{0}, \quad \forall\, i,\ j = 1, 2, \ldots, N,$$

and

$$\mathrm{E}[\mathbf{u}_t (\mathbf{w}_0 - \boldsymbol{\mu}_0)^T] = \mathbf{0}, \qquad \mathrm{E}[\mathbf{v}_t (\mathbf{w}_0 - \boldsymbol{\mu}_0)^T] = \mathbf{0}, \qquad \forall t = 1, 2, \ldots, N.$$

Before turning to the derivation of recursion (4), it is appropriate to discuss some implications arising from the assumptions made regarding the SSM (1), which are essential for the subsequent derivation of the Kalman recursion. First, since $\mathbf{y}_{t-1}$ is a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_{t-1}$, $\mathbf{v}_1, \ldots, \mathbf{v}_{t-1}$, and $\mathbf{w}_0$, and since $\mathbf{v}_t$ is independent of $\mathbf{u}_1, \ldots, \mathbf{u}_{t-1}$, $\mathbf{v}_1, \ldots, \mathbf{v}_{t-1}$ and $\mathbf{w}_0$, it is straightforward that $\mathbf{v}_t$ is independent of $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, i.e., $\mathbf{v}_t$ given $\mathbf{Y}_{t-1}$ equals $\mathbf{v}_t$. Second, since the initial state vector $\mathbf{w}_0$ and the disturbances $\mathbf{u}_1, \ldots, \mathbf{u}_N$, $\mathbf{v}_1, \ldots, \mathbf{v}_N$ are normally distributed, it follows from the linearity of equations (1a) and (1b) that $\mathbf{w}_1, \ldots, \mathbf{w}_N$ and $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are normally distributed as well. Consequently, $\mathbf{w}_t$ and $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ as well as $\mathbf{w}_t$ given $\mathbf{Y}_t$ are also normally distributed for $t = 1, 2, \ldots, N$. This directly follows from a well-known Lemma about the conditional distribution of jointly normally distributed random vectors.

Lemma A.1 *Suppose the random vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are jointly normally distributed with mean vector and variance matrix:*

$$\mathrm{E}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_{\mathrm{x}} \\ \boldsymbol{\mu}_{\mathrm{y}} \end{pmatrix} \quad \textit{and} \quad \mathrm{Var}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathrm{xx}} & \boldsymbol{\Sigma}_{\mathrm{xy}} \\ \boldsymbol{\Sigma}_{\mathrm{yx}} & \boldsymbol{\Sigma}_{\mathrm{yy}} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\mathrm{yy}}$ has rank m. Then the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ is normal with mean vector *and variance matrix:*

$$\mathrm{E}[\mathbf{x}|\mathbf{y}] = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}) \quad \textit{and} \quad \mathrm{Var}[\mathbf{x}|\mathbf{y}] = \boldsymbol{\Sigma}_{\mathbf{xx}} - \boldsymbol{\Sigma}_{\mathbf{xy}}\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1}\boldsymbol{\Sigma}_{\mathbf{xy}}{}^{T}.$$

Proof:

See Durbin and Koopman (2012, pp. 77-78).

□

Lemma A.1 is also a fundamental element of the Kalman filter (KF) in the context of a linear SSM with normally distributed disturbances. Starting in $t = 1$, for each period $t = 1, 2, \ldots, N$ the KF performs two steps:

Prediction step: First we use equation (1b) to obtain the mean vector

$$\begin{aligned}
\mathbf{w}_{t|t-1} &= \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_{t-1}] \\
&= \mathrm{E}[\mathbf{F}\ \mathbf{w}_{t-1} + \mathbf{v}_t|\mathbf{Y}_{t-1}] \\
&= \mathbf{F}\ \mathrm{E}[\mathbf{w}_{t-1}|\mathbf{Y}_{t-1}] + \mathrm{E}[\mathbf{v}_t|\mathbf{Y}_{t-1}] \\
&= \mathbf{F}\ \boldsymbol{\mu}_{t-1},
\end{aligned} \tag{A.1}$$

and variance matrix

$$\begin{aligned}
\mathbf{P}_{t|t-1} &= \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_{t-1}] \\
&= \mathrm{Var}[\mathbf{F}\ \mathbf{w}_{t-1} + \mathbf{v}_t|\mathbf{Y}_{t-1}] \\
&= \mathbf{F}\ \mathrm{Var}[\mathbf{w}_{t-1}|\mathbf{Y}_{t-1}]\ \mathbf{F}^T + \mathrm{Var}[\mathbf{v}_t|\mathbf{Y}_{t-1}] \\
&= \mathbf{F}\ \mathbf{C}_{t-1}\ \mathbf{F}^T + \mathbf{Q},
\end{aligned} \tag{A.2}$$

of $\mathbf{w}_t$ given $\mathbf{Y}_{t-1}$, where $\boldsymbol{\mu}_{t-1}$ and $\mathbf{C}_{t-1}$ are known from a previous iteration or in case of $t = 1$ directly through the initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$.

Updating step: In this second step, we use Lemma A.1 and the new information derived from $\mathbf{y}_t$ to compute $\boldsymbol{\mu}_t = \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_t]$ and $\mathbf{C}_t = \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_t]$. Note that given $\mathbf{Y}_{t-1}$, the random vector

$$\begin{pmatrix} \mathbf{w}_t \\ \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} \mathbf{w}_t \\ \mathbf{h} + \mathbf{H}\mathbf{w}_t + \mathbf{u}_t \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{h} \end{pmatrix} + \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{H} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{w}_t \\ \mathbf{u}_t \end{pmatrix}$$

is jointly normally distributed since it is linear in $\mathbf{w}_t$ given $\mathbf{Y}_{t-1}$ and $\mathbf{u}_t$. Thus $\mathbf{w}_t$ given $\mathbf{Y}_{t-1}$ and $\mathbf{y}_t$ given $\mathbf{Y}_{t-1}$ are jointly normally distributed with mean vector and variance matrix:

$$\mathrm{E}\left(\begin{matrix}\mathbf{w}_t\\ \mathbf{y}_t\end{matrix}\bigg|\mathbf{Y}_{t-1}\right)=\begin{pmatrix}\mathbf{w}_{t|t-1}\\ \mathbf{h}+\mathbf{H}\mathbf{w}_{t|t-1}\end{pmatrix}\quad\text{and}\quad \mathrm{Var}\left(\begin{matrix}\mathbf{w}_t\\ \mathbf{y}_t\end{matrix}\bigg|\mathbf{Y}_{t-1}\right)=\begin{pmatrix}\mathbf{P}_{t|t-1} & \mathbf{X}_t\\ \mathbf{X}_t^T & \mathbf{U}_t\end{pmatrix}.$$

with

$$\begin{aligned}
\mathbf{U}_t &:= \mathrm{Var}[\mathbf{y}_t|\mathbf{Y}_{t-1}] = \mathrm{Var}[\mathbf{h}+\mathbf{H}\,\mathbf{w}_t+\mathbf{u}_t|\mathbf{Y}_{t-1}]\\
&= \mathbf{H}\ \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_{t-1}]\ \mathbf{H}^T+\mathrm{Var}[\mathbf{u}_t|\mathbf{Y}_{t-1}]\\
&= \mathbf{H}\ \mathbf{P}_{t|t-1}\ \mathbf{H}^T+\mathbf{R},\\
\mathbf{X}_t &:= \mathrm{Cov}[\mathbf{w}_t,\mathbf{y}_t|\mathbf{Y}_{t-1}] = \mathrm{E}\left[(\mathbf{w}_t-\mathrm{E}[\mathbf{w}_t|\mathbf{Y}_{t-1}])(\mathbf{y}_t-\mathrm{E}[\mathbf{y}_t|\mathbf{Y}_{t-1}])^T\,\big|\mathbf{Y}_{t-1}\right]\\
&= \mathrm{E}\left[\left(\mathbf{w}_t-\mathbf{w}_{t|t-1}\right)(\mathbf{h}+\mathbf{H}\mathbf{w}_t+\mathbf{u}_t-\mathrm{E}[\mathbf{h}+\mathbf{H}\mathbf{w}_t+\mathbf{u}_t|\mathbf{Y}_{t-1}])^T\,\big|\mathbf{Y}_{t-1}\right]\\
&= \mathrm{E}\left[\left(\mathbf{w}_t-\mathbf{w}_{t|t-1}\right)\left(\mathbf{w}_t-\mathbf{w}_{t|t-1}\right)^T\big|\mathbf{Y}_{t-1}\right]\mathbf{H}^T+\mathrm{E}\left[\left(\mathbf{w}_t-\mathbf{w}_{t|t-1}\right)\mathbf{u}_t^T\big|\mathbf{Y}_{t-1}\right]\\
&= \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_{t-1}]\,\mathbf{H}^T+\mathrm{Cov}[\mathbf{w}_t,\mathbf{u}_t|\mathbf{Y}_{t-1}]\\
&= \mathbf{P}_{t|t-1}\ \mathbf{H}^T.
\end{aligned}$$

Hence, the mean vector $\boldsymbol{\mu}_t$ and the variance matrix $\mathbf{C}_t$ of $\mathbf{w}_t$ given $\mathbf{Y}_t$ follow directly from Lemma A.1 as

$$\begin{aligned}
\boldsymbol{\mu}_t &= \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_t] = \mathrm{E}[\mathbf{w}_t|\mathbf{y}_t,\mathbf{Y}_{t-1}]\\
&= \mathbf{w}_{t|t-1}+\mathbf{X}_t\ \mathbf{U}_t^{-1}(\mathbf{y}_t-\mathrm{E}[\mathbf{y}_t|\mathbf{Y}_{t-1}])\\
&= \mathbf{w}_{t|t-1}+\mathbf{P}_{t|t-1}\mathbf{H}^T\ \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T+\mathbf{R}\right]^{-1}\left(\mathbf{y}_t-\mathbf{h}-\mathbf{H}\mathbf{w}_{t|t-1}\right) \qquad \text{(A.3)}\\
\mathbf{C}_t &= \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_t] = \mathrm{Var}[\mathbf{w}_t|\mathbf{y}_t,\mathbf{Y}_{t-1}]\\
&= \mathbf{P}_{t|t-1}-\mathbf{X}_t\ \mathbf{U}_t^{-1}\ \mathbf{X}_t^T\\
&= \mathbf{P}_{t|t-1}-\mathbf{P}_{t|t-1}\mathbf{H}^T\ \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T+\mathbf{R}\right]^{-1}\ \mathbf{H}\mathbf{P}_{t|t-1}. \qquad \text{(A.4)}
\end{aligned}$$

If we define $\mathbf{K}_t$ for all $t=1,2,\ldots,N$ as in (3) and replace $\mathbf{y}_t-\mathbf{h}$ with $\mathbf{y}_t^{(\mathbf{h})}$, the Kalman recursion (4) follows directly from (A.1) – (A.4).

B Convergence properties of the Riccati difference equation

The purpose of this appendix is to give the reader a general idea under which conditions the Riccati difference equations (RDEs)

$$\begin{aligned}
\mathbf{C}_t &= \mathbf{F}\mathbf{C}_{t-1}\mathbf{F}^T+\mathbf{Q}-\left(\mathbf{F}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T+\bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T+\bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\mathbf{C}_{t-1}\mathbf{F}^T+\bar{\mathbf{G}}^T\right),\\
\mathbf{P}_{t+1|t} &= \mathbf{F}\mathbf{P}_{t|t-1}\mathbf{F}^T-\mathbf{F}\mathbf{P}_{t|t-1}\mathbf{H}^T\ \left[\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T+\mathbf{R}\right]^{-1}\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{F}^T+\mathbf{Q},
\end{aligned}$$

with $\bar{\mathbf{H}} := \mathbf{HF}$, $\bar{\mathbf{G}} := \mathbf{QH}^T$, and $\bar{\mathbf{R}} := \mathbf{HQH}^T + \mathbf{R}$, described by equations (8a) and (8b), have a fixed-point and for which initialization they converge to this fixed-point. To this end, the first part of this appendix deals with cases where either the matrix $\mathbf{R}$ or at least the matrix $\bar{\mathbf{R}}$ is non-singular, drawing from the convergence results provided by de Souza et al. (1986).[20] Since the results of de Souza et al. (1986) are fairly general, we shall also discuss some more frequently consulted conditions — namely *stability*, *observability* and *reachability*, *detectability* and *stabilizability* — sufficient for convergence of RDEs, such as (8a) or (8b). The second part of this appendix contains the formal proof of Proposition 2.1.

B.1 Results by de Souza et al. (1986)

De Souza et al. (1986) provide some general convergence results related to (ordinary) RDEs of the form

$$\boldsymbol{\Sigma}_t = \mathbf{F}\boldsymbol{\Sigma}_{t-1}\mathbf{F}^T - \mathbf{F}\boldsymbol{\Sigma}_{t-1}\mathbf{H}^T \left[\mathbf{H}\boldsymbol{\Sigma}_{t-1}\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{H}\boldsymbol{\Sigma}_{t-1}\mathbf{F}^T + \mathbf{Q}, \quad \forall t = 1, 2, \ldots, N, \tag{B.1}$$

with $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$, $\mathbf{Q} \in \mathbb{R}^{n\times n}$, and $\mathbf{R} \in \mathbb{R}^{m\times m}$. Further, they assume that $\mathbf{Q}$ and $\mathbf{R}$ are symmetric matrices with $\mathbf{Q} = \mathbf{DD}^T \geq \mathbf{0}$, $\mathbf{D} \in \mathbb{R}^{n\times n}$ and $\mathbf{R} > \mathbf{0}$.[21] Thus, their results are directly transferable to RDE (8b) if the variance matrix $\mathbf{R}$ of the measurement error $\mathbf{u}_t$ is non-singular. On the other hand, if $\mathbf{R}$ is singular, we can use their results to study the convergence properties of RDE (8a), which is sometimes called a generalized RDE, provided at least $\bar{\mathbf{R}}$ is non-singular. To do so, we can transform the general RDE (8a) into an ordinary RDE of the form (B.1) using the following lemma:

Lemma B.1 *Suppose the matrices $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\bar{\mathbf{H}} \in \mathbb{R}^{m\times n}$ and $\bar{\mathbf{G}} \in \mathbb{R}^{n\times m}$, as well as the variances matrices $\mathbf{Q} \in \mathbb{R}^{n\times n}$ and $\bar{\mathbf{R}} \in \mathbb{R}^{m\times m}$ with $\bar{\mathbf{R}} > \mathbf{0}$ refer to the generalized RDE*

$$\bar{\boldsymbol{\Sigma}}_t = \mathbf{F}\bar{\boldsymbol{\Sigma}}_{t-1}\mathbf{F}^T + \mathbf{Q} - \left(\mathbf{F}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\mathbf{F}^T + \bar{\mathbf{G}}^T\right), \tag{B.2}$$

that generates the matrix sequence $\{\bar{\boldsymbol{\Sigma}}_t\}_{t=0}^N$ with a variance matrix $\bar{\boldsymbol{\Sigma}}_0 \geq \mathbf{0}$. Then the RDE

$$\bar{\boldsymbol{\Sigma}}_t = \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T,$$

with $\bar{\mathbf{F}} := \mathbf{F} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}$ and $\bar{\mathbf{Q}} := \mathbf{Q} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T$, is equivalent to (B.2).

Proof:
The statement follows from the fact that we may rewrite the right-hand side of (B.2) to

$$\mathbf{F}\bar{\boldsymbol{\Sigma}}_{t-1}\mathbf{F}^T + \mathbf{Q} - \left(\mathbf{F}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\mathbf{F}^T + \bar{\mathbf{G}}^T\right)$$

[20] Note that the non-singularity of $\mathbf{R}$ implies that $\bar{\mathbf{R}}$ must be also non-singular.
[21] Note that the notation $\mathbf{A} > 0$ (or $\mathbf{A} \geq 0$) means that the matrix $\mathbf{A}$ is positive-definite (or positive-semi-definite).

$$
\begin{aligned}
&= \left(\bar{\mathbf{F}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\right) \bar{\boldsymbol{\Sigma}}_{t-1} \left(\bar{\mathbf{F}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\right)^T + \bar{\mathbf{Q}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \left[\left(\bar{\mathbf{F}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\right) \bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1} \left(\bar{\mathbf{F}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\right)^T + \bar{\mathbf{G}}^T\right] \\
&= \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{Q}} + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \left[\bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}^T\right] \\
&= \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T \\
&\quad + \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}^T\right] \\
&\quad - \left[\bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T \\
&\quad - \left[\bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{G}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}^T\right] \\
&= \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T \\
&\quad + \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T \\
&\quad - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right] \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&= \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T \\
&\quad + \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&\quad - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T \\
&= \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T \left[\bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1} \bar{\mathbf{H}}\bar{\boldsymbol{\Sigma}}_{t-1}\bar{\mathbf{F}}^T.
\end{aligned}
$$

□

Stabilizing and strong solutions: Note that if the RDE (B.1) converges to a fixed matrix $\boldsymbol{\Sigma}$, we may state that $\boldsymbol{\Sigma}$ is a solution to the discrete algebraic Riccati equation (DARE)

$$
\boldsymbol{\Sigma} = \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^T - \mathbf{F}\boldsymbol{\Sigma}\mathbf{H}^T \left[\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{H}\boldsymbol{\Sigma}\mathbf{F}^T + \mathbf{Q}. \tag{B.3}
$$

We call (B.3) the DARE corresponding to the RDE (B.1). Considering the convergence of (B.1), two types of solutions to the DARE (B.3) are of particular importance.

Definition B.1 *Suppose the matrices* $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$, $\mathbf{Q} \in \mathbb{R}^{n\times n}$ *and* $\mathbf{R} \in \mathbb{R}^{m\times m}$ *refer to the RDE* (B.1) *that generates the matrix sequence* $\{\boldsymbol{\Sigma}_t\}_{t=0}^N$ *with a variance matrix* $\boldsymbol{\Sigma}_0 \geq \mathbf{0}$. *Further suppose that* (B.3) *is the DARE corresponding to the RDE* (B.1)*, then a real symmetric matrix* $\boldsymbol{\Sigma} \geq \mathbf{0}$ *satisfying the DARE* (B.3) *is called a* ***stabilizing / strong*** *solution, if and only if the eigenvalues of the matrix*

$$
\tilde{\mathbf{F}} = \mathbf{F}\left(\mathbf{I} - \boldsymbol{\Sigma}\mathbf{H}^T \left[\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{H}\right)
$$

*are **inside / inside or on** the unit circle.*[22]

As we will see, (B.1) often converges towards its strong (stabilizing) solution, provided this solution exists. Furthermore, to show that the DAREs

$$\mathbf{C}_{+} = \mathbf{F}\mathbf{C}_{+}\mathbf{F}^{T} + \mathbf{Q} - \left(\mathbf{F}\mathbf{C}_{+}\bar{\mathbf{H}}^{T} + \bar{\mathbf{G}}\right)\left[\bar{\mathbf{H}}\mathbf{C}_{+}\bar{\mathbf{H}}^{T} + \bar{\mathbf{R}}\right]^{-1}\left(\bar{\mathbf{H}}\mathbf{C}_{+}\mathbf{F}^{T} + \bar{\mathbf{G}}^{T}\right), \tag{B.4}$$

$$\mathbf{P}_{+} = \mathbf{F}\mathbf{P}_{+}\mathbf{F}^{T} - \mathbf{F}\mathbf{P}_{+}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{P}_{+}\mathbf{F}^{T} + \mathbf{Q}, \tag{B.5}$$

corresponding to the RDEs (8a) and (8b) have strong (stabilizing) solutions, it is sufficient to show that one of the DAREs (B.4) and (B.5) has a strong (stabilizing) solution. To see this we propose the following Lemma:

Lemma B.2 *The matrix* $\mathbf{C}_{+}$ *is a **stabilizing / strong** solution to* (B.4)*, if and only if* $\mathbf{P}_{+}$ *is a **strong / stabilizing** solution to* (B.5).

Proof:
From Definition B.1 follows that $\mathbf{P}_{+}$ is a **stabilizing** / **strong** solution to (B.5), if and only if the eigenvalues of the matrix

$$\tilde{\mathbf{F}}_{\mathbf{P}_{+}} = \mathbf{F}\left(\mathbf{I} - \mathbf{P}_{+}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\right)$$

are **inside** / **inside or on** the unit circle. Analogously, using Lemma B.1 we may state that $\mathbf{C}_{+}$ is a **stabilizing** / **strong** solution to (B.4), if and only if the eigenvalues of the matrix

$$\tilde{\mathbf{F}}_{\mathbf{C}_{+}} = \bar{\mathbf{F}}\left(\mathbf{I} - \mathbf{C}_{+}\bar{\mathbf{H}}^{T}\left[\bar{\mathbf{H}}\mathbf{C}_{+}\bar{\mathbf{H}}^{T} + \bar{\mathbf{R}}\right]^{-1}\bar{\mathbf{H}}\right)$$

are **inside** / **inside or on** the unit circle. Thus, to prove the claim of Lemma B.2 we will show that $\tilde{\mathbf{F}}_{\mathbf{P}_{+}}$ and $\tilde{\mathbf{F}}_{\mathbf{C}_{+}}$ share the same set of eigenvalues. To see this, note that using the definitions of $\bar{\mathbf{F}}$, $\bar{\mathbf{H}}$, and $\bar{\mathbf{R}}$ as well as (4b), we may write

$$\begin{aligned}
\tilde{\mathbf{F}}_{\mathbf{C}_{+}} &= \bar{\mathbf{F}}\left(\mathbf{I} - \mathbf{C}_{+}\bar{\mathbf{H}}^{T}\left[\bar{\mathbf{H}}\mathbf{C}_{+}\bar{\mathbf{H}}^{T} + \bar{\mathbf{R}}\right]^{-1}\bar{\mathbf{H}}\right)\\
&= \left(\mathbf{F} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}\right)\left(\mathbf{I} - \mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{F}\mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T} + \mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\\
&= \left(\mathbf{F} - \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\left(\mathbf{I} - \mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{F}\mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T} + \mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\\
&= \left(\mathbf{F} - \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\left(\mathbf{I} - \mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\\
&= \left(\mathbf{I} - \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\right)\left(\mathbf{F} - \mathbf{F}\mathbf{C}_{+}\mathbf{F}^{T}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\\
&= \left(\mathbf{I} - \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\right)\left(\mathbf{F} - \mathbf{P}_{+}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F} + \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}\right)\\
&= \mathbf{F} - \mathbf{P}_{+}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F} + \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{P}_{+}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F} - \mathbf{Q}\mathbf{H}^{T}\left[\mathbf{H}\mathbf{Q}\mathbf{H}^{T} + \mathbf{R}\right]^{-1}\mathbf{H}\mathbf{F}
\end{aligned}$$

[22] This definition is taken from Chan et al. (1984) and de Souza et al. (1986).

$$
\begin{aligned}
&+ \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HP}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&- \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HQH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
= \; & \mathbf{F} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} + \mathbf{QH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} - \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&+ \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HP}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&- \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HQH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&- \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{R}\left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&+ \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{R}\left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
= \; & \mathbf{F} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} + \mathbf{QH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} - \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&+ \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]\left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&- \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \left[\mathbf{HQH}^T + \mathbf{R}\right]\left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
= \; & \mathbf{F} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} + \mathbf{QH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} - \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
&+ \mathbf{QH}^T \left[\mathbf{HQH}^T + \mathbf{R}\right]^{-1} \mathbf{HF} - \mathbf{QH}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
= \; & \mathbf{F} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{HF} \\
= \; & \left(\mathbf{I} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{H}\right)\mathbf{F}
\end{aligned}
$$

Since $\mathbf{F}$ and $\left(\mathbf{I} - \mathbf{P}_+\mathbf{H}^T \left[\mathbf{HP}_+\mathbf{H}^T + \mathbf{R}\right]^{-1} \mathbf{H}\right)$ are both square matrices, the matrix $\tilde{\mathbf{F}}_{\mathbf{C}_+}$ must have the same set of eigenvalues as the matrix $\tilde{\mathbf{F}}_{\mathbf{P}_+}$.[23] This completes the proof.

□

Some concepts of linear systems theory: To eventually obtain conditions under which RDE (B.1) converges to a fixed matrix $\boldsymbol{\Sigma}$, we introduce the concepts of stability, observability, and reachability from linear system theory:[24]

Definition B.2 *Suppose the matrices* $\mathbf{F} \in \mathbb{R}^{n\times n}$*,* $\mathbf{H} \in \mathbb{R}^{m\times n}$*,* $\mathbf{R} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{Q} \in \mathbb{R}^{n\times n}$*, with* $\mathbf{Q} = \mathbf{DD}^T$*,* $\mathbf{D} \in \mathbb{R}^{n\times n}$*, refer to the RDE* (B.1) *that generates the matrix sequence* $\{\boldsymbol{\Sigma}_t\}_{t=0}^N$ *with a variance matrix* $\boldsymbol{\Sigma}_0 \geq \mathbf{0}$*. Further suppose that* (B.3) *is the DARE corresponding to the RDE* (B.1)*, then*

(i) The matrix $\mathbf{F}$ *is called stable, if and only if for any eigenvalue* λ *of the matrix* $\mathbf{F}$ *it holds that*

$$|\lambda| < 1.$$

(ii) The pair $(\mathbf{H}, \mathbf{F})$ *is called observable, if and only if*

$$\mathrm{rk}\begin{pmatrix}\mathbf{H}^T & \mathbf{F}^T\,\mathbf{H}^T & \cdots & (\mathbf{F}^T)^{n-1}\,\mathbf{H}^T\end{pmatrix} = n.$$

[23] Note that from $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$ follows that the matrices $\mathbf{C} = \mathbf{AB}$ and $\mathbf{D} = \mathbf{BA}$ share the same set of eigenvalues. See e.g. Theorem 6.12 by Searle and Khuri (2017, pp. 140).

[24] See also (Gu, 2012, Chapter 3).

(iii) The pair $(\mathbf{F}, \mathbf{D})$ *is called reachable, if and only if*

$$\text{rk}\begin{pmatrix}\mathbf{D} & \mathbf{F}\,\mathbf{D} & \cdots & \mathbf{F}^{n-1}\,\mathbf{D}\end{pmatrix} = n.$$

With respect to the SSM (1), a stable transition matrix $\mathbf{F}$ ensures that $\mathbf{w}_t$ follows a stationary process so that the unforced system

$$\mathbf{w}_t = \mathbf{F}\mathbf{w}_{t-1}, \quad \mathbf{w}_0 \neq 0,$$

is asymptotically stable, i.e., $\lim_{t\to\infty} \mathbf{w}_t = \mathbf{0}$. Observability and reachability are dual concepts, i.e., for an observable pair $(\mathbf{H}, \mathbf{F})$, we may claim that the pair $(\mathbf{F}^T, \mathbf{H}^T)$ is reachable and vice versa. Observability of the pair $(\mathbf{H}, \mathbf{F})$ can also be understood in the sense that, in the case of an unforced system

$$\mathbf{y}_t = \mathbf{H}\mathbf{w}_t, \quad \mathbf{w}_t = \mathbf{F}\mathbf{w}_{t-1},$$

there is some $l \in \mathbb{N}$, such that the initial state vector $\mathbf{w}_0$ may be obtained from $\{\mathbf{y}_t\}_{t=1}^{l}$ (see Gu, 2012, pp. 70). Reachability of the pair $(\mathbf{F}, \mathbf{D})$, on the other hand, can be interpreted in the sense that there is a bounded control input $\{\mathbf{v}_t\}_{t=1}^{l}$, $l \in \mathbb{N}$, so that the system

$$\mathbf{w}_t = \mathbf{F}\mathbf{w}_{t-1} + \mathbf{D}\mathbf{v}_t,$$

can reach a state $\mathbf{w}^*$, i.e., $\mathbf{w}_l = \mathbf{w}^*$, for a given initial state $\mathbf{w}_0$ (see Gu, 2012, pp. 75). As we will see, observability and reachability are sufficient conditions for (B.1) to converge to a stabilizing solution for $\boldsymbol{\Sigma}_0 > \mathbf{0}$. However, since especially the assumption of reachability does not hold for a variety of econometric models,[25] somewhat weaker concepts than observability and reachablility are detectability and stabilizability:

Definition B.3 *Suppose the matrices* $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$, $\mathbf{R} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{Q} \in \mathbb{R}^{n\times n}$, *with* $\mathbf{Q} = \mathbf{D}\mathbf{D}^T$, $\mathbf{D} \in \mathbb{R}^{n\times n}$, *refer to the RDE* (B.1) *that generates the matrix sequence* $\{\boldsymbol{\Sigma}_t\}_{t=0}^{N}$ *with a variance matrix* $\boldsymbol{\Sigma}_0 \geq \mathbf{0}$. *Further suppose that* (B.3) *is the DARE corresponding to the RDE* (B.1)*, then*

(i) The pair $(\mathbf{H}, \mathbf{F})$ *is called detectable, if for any eigenvalue* λ *of the matrix* $\mathbf{F}$ *with* $|\lambda| \geq 1$, *there does not exist a n-dimensional eigenvector* $\mathbf{q} \neq \mathbf{0}$ *such that*

$$\mathbf{F}\mathbf{q} = \lambda\mathbf{q}, \quad \mathbf{H}\mathbf{q} = \mathbf{0}.$$

[25] See e.g. Harvey (1990b, pp. 118) who illustrates that a non-invertible moving average process of order one will always be observable, but never be reachable.

(ii) *The pair* $(\mathbf{F}, \mathbf{D})$ *is called stabilizable, if for any eigenvalue* λ *of the matrix* $\mathbf{F}^T$ *with* $|\lambda| \geq 1$*, there does not exist a n-dimensional eigenvector* $\mathbf{q} \neq \mathbf{0}$ *such that*

$$\mathbf{F}^T \mathbf{q} = \lambda \mathbf{q}, \quad \mathbf{D}^T \mathbf{q} = \mathbf{0}.$$

(iii) *Further an eigenvalue* λ *of the matrix* $\mathbf{F}$ *is said to be* $(\mathbf{F}, \mathbf{D})$*-unreachable (of rank p) if and only if there exists a set of (p) n-dimensional generalized eigenvectors* $\mathbf{q}_i \neq 0$ *with* $i = 1, \ldots, p$ *and* $\mathbf{q}_0 = 0$ *such that*

$$\mathbf{F}^T \mathbf{q}_i = \lambda \mathbf{q}_i + \mathbf{q}_{i-1}, \quad \mathbf{D}^T \mathbf{q}_i = 0.$$

In the following lemma, we postulate some well-known links from detectability and stabilizability to the concepts in Definition B.2:

Lemma B.3 *Suppose the matrices* $\mathbf{F} \in \mathbb{R}^{n \times n}$*,* $\mathbf{H} \in \mathbb{R}^{m \times n}$*,* $\mathbf{R} \in \mathbb{R}^{m \times m}$ *and* $\mathbf{Q} \in \mathbb{R}^{n \times n}$*, with* $\mathbf{Q} = \mathbf{D}\mathbf{D}^T$*,* $\mathbf{D} \in \mathbb{R}^{n \times n}$*, refer to the RDE* (B.1) *that generates the matrix sequence* $\{\mathbf{\Sigma}_t\}_{t=0}^N$ *with a variance matrix* $\mathbf{\Sigma}_0 \geq \mathbf{0}$*. Further suppose that* (B.3) *is the DARE corresponding to the RDE* (B.1)*, then we may state that:*

(i) *If the matrix* $\mathbf{F}$ *is stable, then the pair* $(\mathbf{H}, \mathbf{F})$ *is detectable.*

(ii) *If the matrix* $\mathbf{F}$ *is stable, then the pair* $(\mathbf{F}, \mathbf{D})$ *is stabilizable.*

(iii) *If the pair* $(\mathbf{H}, \mathbf{F})$ *is observable it is also detectable.*

(iv) *If the pair* $(\mathbf{F}, \mathbf{D})$ *is reachable it is also stabilizable.*

(v) *The pair* $(\mathbf{F}, \mathbf{D})$ *is stabilizable, if and only if the matrix* $\mathbf{F}$ *has no* $(\mathbf{F}, \mathbf{D})$*-unreachable eigenvalues on or outside the unit circle, i.e.* $\lambda \leq 1$*.*

(vi) *The pair is* $(\mathbf{F}, \mathbf{D})$ *is reachable, if and only if the matrix* $\mathbf{F}$ *has no* $(\mathbf{F}, \mathbf{D})$*-unreachable eigenvalues.*

Proof:

(i) Note that the stability of $\mathbf{F}$ implies that there are no eigenvalues λ of the matrix $\mathbf{F}$ with $|\lambda| \geq 1$. Consequently, for all eigenvalues with $|\lambda| \geq 1$ (where there are none), there does not exist an n-dimensional vector $\mathbf{q} \neq \mathbf{0}$ such that

$$\mathbf{F}\mathbf{q} = \lambda \mathbf{q}, \quad \mathbf{H}\mathbf{q} = \mathbf{0}.$$

(ii) Note that the stability of $\mathbf{F}$ implies that there are no eigenvalues λ of the matrix $\mathbf{F}^T$ with $|\lambda| \geq 1$. Consequently, for all eigenvalues with $|\lambda| \geq 1$ (where there are none), there does not exist an n-dimensional vector $\mathbf{q} \neq \mathbf{0}$ such that

$$\mathbf{F}^T\, \mathbf{q} = \lambda \mathbf{q}, \quad \mathbf{D}^T\, \mathbf{q} = \mathbf{0}.$$

(iii) The statement follows from the fact that observability of the pair $(\mathbf{H}, \mathbf{F})$ implies that the matrix $\mathbf{F}$ has no eigenvector $\mathbf{q}$ (corresponding to an eigenvalue λ) such that

$$\mathbf{F}\mathbf{q} = \lambda \mathbf{q}, \quad \mathbf{H}\mathbf{q} = \mathbf{0}.$$

This is true, since otherwise we could write

$$\begin{aligned}
\mathbf{q}^T \begin{pmatrix} \mathbf{H}^T & \mathbf{F}^T\,\mathbf{H}^T & \cdots & (\mathbf{F}^T)^{n-1}\,\mathbf{H}^T \end{pmatrix} &= \begin{pmatrix} \mathbf{q}^T\,\mathbf{H}^T & \mathbf{q}^T\,\mathbf{F}^T\,\mathbf{H}^T & \cdots & \mathbf{q}^T\,(\mathbf{F}^T)^{n-1}\,\mathbf{H}^T \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{H}\,\mathbf{q})^T & (\mathbf{H}\,\mathbf{F}\,\mathbf{q})^T & \cdots & (\mathbf{H}\,\mathbf{F}^{n-1}\,\mathbf{q})^T \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{H}\,\mathbf{q})^T & \lambda\,(\mathbf{H}\,\mathbf{q})^T & \cdots & \lambda^{n-1}\,(\mathbf{H}\,\mathbf{q})^T \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \\
&= \mathbf{0},
\end{aligned}$$

so that

$$\mathrm{rk}\begin{pmatrix} \mathbf{H}^T & \mathbf{F}^T\,\mathbf{H}^T & \cdots & (\mathbf{F}^T)^{n-1}\,\mathbf{H}^T \end{pmatrix} < n.$$

(iv) The statement follows from the fact that reachability of the pair $(\mathbf{F}, \mathbf{D})$ implies that the matrix $\mathbf{F}^T$ has no eigenvector $\mathbf{q}$ (corresponding to an eigenvalue λ) such that

$$\mathbf{F}^T\mathbf{q} = \lambda \mathbf{q}, \quad \mathbf{D}^T\mathbf{q} = \mathbf{0}.$$

This is true, since otherwise we could write

$$\begin{aligned}
\mathbf{q}^T \begin{pmatrix} \mathbf{D} & \mathbf{F}\,\mathbf{D} & \cdots & \mathbf{F}^{n-1}\,\mathbf{D} \end{pmatrix} &= \begin{pmatrix} \mathbf{q}^T\,\mathbf{D} & \mathbf{q}^T\,\mathbf{F}\,\mathbf{D} & \cdots & \mathbf{q}^T\,\mathbf{F}^{n-1}\,\mathbf{H} \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{D}^T\,\mathbf{q})^T & (\mathbf{D}^T\,\mathbf{F}^T\,\mathbf{q})^T & \cdots & (\mathbf{D}^T\,(\mathbf{F}^T)^{n-1}\,\mathbf{q})^T \end{pmatrix} \\
&= \begin{pmatrix} (\mathbf{D}^T\,\mathbf{q})^T & \lambda\,(\mathbf{D}^T\,\mathbf{q})^T & \cdots & \lambda^{n-1}\,(\mathbf{D}^T\,\mathbf{q})^T \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \\
&= \mathbf{0},
\end{aligned}$$

so that

$$\mathrm{rk}\begin{pmatrix} \mathbf{D} & \mathbf{F}\,\mathbf{D} & \cdots & \mathbf{F}^{n-1}\,\mathbf{D} \end{pmatrix} < n.$$

(v) Follows directly from Definition B.3 (ii)-(iii).

(vi) Follows from Theorem 3.10. Gu (2012, pp. 77).

□

Roughly speaking, we may describe **detectability / stabilizability** as the claim that all parts, or more precisely all eigenvalues of the transition matrix $\mathbf{F}$, are either $(\mathbf{H},\mathbf{F})$ **observable** / $(\mathbf{F},\mathbf{D})$ **reachable** or stable. Please note that some authors, e.g., Harvey (1990b, pp. 115), use the related concept of Controllability instead of the concept of Reachability. For more details on linear systems theory, we refer the reader to Gu (2012, Chapter 3) as well as Anderson and Moore (1979, Appendix C).

Some general convergence results: Finally, we collect the main results on the convergence of the RDE (B.1) provided by de Souza et al. (1986) in the following Proposition:

Proposition B.1 *Suppose the matrices* $\mathbf{F} \in \mathbb{R}^{n\times n}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$, $\mathbf{R} \in \mathbb{R}^{m\times m}$ *and* $\mathbf{Q} \in \mathbb{R}^{n\times n}$, *with* $\mathbf{Q} = \mathbf{D}\mathbf{D}^T$, $\mathbf{D} \in \mathbb{R}^{n\times n}$, *refer to the RDE* (B.1) *that generates the matrix sequence* $\{\boldsymbol{\Sigma}_t\}_{t=0}^N$ *with a variance matrix* $\boldsymbol{\Sigma}_0 \geq \mathbf{0}$. *Further suppose that* (B.3) *is the DARE corresponding to the RDE* (B.1)*, then*

(i) The strong solution $\boldsymbol{\Sigma}_s$ *of the DARE* (B.3) *exists and is unique if and only if* $(\mathbf{H},\mathbf{F})$ *is detectable. Furthermore, subject to* $(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_s) \geq \mathbf{0}$, *the RDE* (B.1) *converges to the strong solution (i.e.* $\lim_{t\to\infty} \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_s$*).*

(ii) The strong solution $\boldsymbol{\Sigma}_s$ *is the only non-negative definite solution of the DARE* (B.3) *if and only if* $(\mathbf{H},\mathbf{F})$ *is detectable and* $\mathbf{F}$ *has no* $(\mathbf{F},\mathbf{D})$*-unreachable eigenvalues outside the unit circle.*

(iii) The strong solution $\boldsymbol{\Sigma}_s$ *is a stabilizing solution of the DARE* (B.3) *if and only if* $(\mathbf{H},\mathbf{F})$ *is detectable and* $\mathbf{F}$ *has no* $(\mathbf{F},\mathbf{D})$*-unreachable eigenvalues on the unit circle. Furthermore, subject to* $\boldsymbol{\Sigma}_0 > \mathbf{0}$, *the RDE* (B.1) *converges to the strong and stabilizing solution* $\boldsymbol{\Sigma}_s$*.*

Proof:

(i) See de Souza et al. (1986, Theorem 3.2-A and theorem 4.2).

(ii) See de Souza et al. (1986, Theorem 3.2-B).

(iii) See de Souza et al. (1986, Theorem 3.2-C and theorem 4.1).

□

From Proposition B.1(i)-(iii) and Lemma B.3(v) follows the well-known result that the RDE (B.1) converges to a stabilizing solution if the pair $(\mathbf{H},\mathbf{F})$ is detectable while the pair $(\mathbf{F},\mathbf{D})$ is stabilizable. Furthermore, it follows from Lemma B.3(i)-(iii) that the same holds true if $\mathbf{F}$ is stable and/or if the pair $(\mathbf{H},\mathbf{F})$ is observable while the pair $(\mathbf{F},\mathbf{D})$ is reachable. However, Proposition B.1(i) also allows to investigate the existence and convergence to a strong solution.

B.2 Proof of Proposition 2.1

To prove claims (i) and (ii) of Proposition 2.1, we will consult the results of Proposition B.1 and Lemmas B.1, B.2 and B.3, while the proof of claim (iii) basis on Proposition 13.1 by (Hamilton, 1994, pp. 390) and Lemma B.2.

Statement (i): To prove the claim (i), we first consider the sequence $\{\mathbf{P}_{t|t-1}\}_{t=1}^{N}$ determined by the (ordinary) RDE (8b). Since $\mathbf{R}$ is a non-singular matrix by assumption, we may use the results by de Souza et al. (1986) to analyze the convergence behavior of $\{\mathbf{P}_{t|t-1}\}_{t=1}^{N}$. Using the fact that the matrix $\mathbf{F}$ is stable by assumption, it follows from Lemma B.3 (i), (ii), and (v) that $(\mathbf{H},\mathbf{F})$ is detectable and $\mathbf{F}$ has no $(\mathbf{F},\mathbf{D})$-unreachable eigenvalues on (or outside) the unit circle. Thus, the claim related to the sequence $\{\mathbf{P}_{t|t-1}\}_{t=1}^{N}$ follows directly from the „if " part of Proposition B.1 (iii). The claim related to the sequence $\{\mathbf{C}_t\}_{t=1}^{N}$ then follows directly from Lemma B.2 and the „only if" part of Proposition B.1 (iii).

Statement (ii): To prove the claim (ii), we first consider the sequence $\{\mathbf{C}_t\}_{t=1}^{N}$ determined by the (general) RDE (8a). Since $\bar{\mathbf{R}}$ is a non-singular matrix by assumption, we may use Lemma B.1 to transform (8a) into the (ordinary) RDE

$$\mathbf{C}_t = \bar{\mathbf{F}}\mathbf{C}_{t-1}\bar{\mathbf{F}}^T + \bar{\mathbf{Q}} - \bar{\mathbf{F}}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T\left[\bar{\mathbf{H}}\mathbf{C}_{t-1}\bar{\mathbf{H}}^T + \bar{\mathbf{R}}\right]^{-1}\bar{\mathbf{H}}\mathbf{C}_{t-1}\bar{\mathbf{F}}^T, \tag{B.6}$$

with $\bar{\mathbf{F}} := \mathbf{F} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{H}}$ and $\bar{\mathbf{Q}} := \mathbf{Q} - \bar{\mathbf{G}}\bar{\mathbf{R}}^{-1}\bar{\mathbf{G}}^T$. Based on (B.6), we may use the results by de Souza et al. (1986) to analyze the convergence behavior of $\{\mathbf{C}_t\}_{t=1}^{N}$. Using the fact that the matrix $\bar{\mathbf{F}}$ is stable by assumption, it follows from Lemma B.3 (i), (ii), and (v) that $(\mathbf{H},\mathbf{F})$ is detectable and $\bar{\mathbf{F}}$ has no $(\bar{\mathbf{F}},\bar{\mathbf{D}})$-unreachable eigenvalues on (or outside) the unit circle. Thus, the claim related to the sequence $\{\mathbf{C}_t\}_{t=1}^{N}$ follows directly from the „if " part of Proposition B.1 (iii). The claim related to the sequence $\{\mathbf{P}_{t|t-1}\}_{t=1}^{N}$ then follows directly from Lemma B.2 and the „only if " part of Proposition B.1 (iii).

Statement (iii): Hamilton (1994, Chapter 13) shows that using the unconditional initialization, the sequence $\{\mathbf{P}_{t|t-1}\}_{t=1}^{N}$ is non-increasing, i.e., $\mathbf{P}_{t|t-1} - \mathbf{P}_{t+1|t}$ is positive semi-definite for all $t = 1, \dots, N-1$, and converges to a strong solution $\mathbf{P}_+$, with $\mathbf{P}_{1|0} - \mathbf{P}_+ = \mathbf{C}_0 - \mathbf{P}_+ \geq \mathbf{0}$.[26] Furthermore, it follows from (3) and (4f) that

$$\mathbf{P}_+ - \mathbf{C}_+ = \mathbf{P}_+\mathbf{H}^T\left(\mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}\right)^{-1}\left(\mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}\right)\left(\mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{H}\mathbf{P}_+ \geq \mathbf{0}$$

so that

$$\mathbf{C}_0 - \mathbf{C}_+ \geq \mathbf{C}_0 - \mathbf{P}_+ = \mathbf{P}_{1|0} - \mathbf{P}_+ \geq \mathbf{0}.$$

Since we know from Lemma B.2 that $\mathbf{C}_+$ is a strong solution of RDE (8a), the claim related to the sequence $\{\mathbf{C}_t\}_{t=1}^N$ follows directly from Proposition B.1 (i).

□

C Derivation of the steady-state Kalman filter

In this appendix, we provide the formal derivation of the steady-state Kalman filter (SKF) (11) and the steady-state log-likelihood (12). To do so, note that if we initialize the KF at $(\boldsymbol{\mu}_{0,+}, \mathbf{C}_+)$, where $\mathbf{C}_+$ is a solution to (B.4), $\mathbf{C}_t = \mathbf{C}_+$ for all $t = 1, 2, \dots, N$. Furthermore, the quantities $\mathbf{P}_{t|t-1}$, $\mathbf{U}_t$, and $\mathbf{K}_t$ become time-invariant, too. If we denote their steady-state equivalents as $\mathbf{P}_+$, $\mathbf{U}_+$, and $\mathbf{K}_+$, it follows directly from (4b), (4d), and (3) that

$$\mathbf{P}_+ = \mathbf{F}\mathbf{C}_+\mathbf{F}^T + \mathbf{Q}, \qquad \mathbf{U}_+ = \mathbf{H}\mathbf{P}_+\mathbf{H}^T + \mathbf{R}, \qquad \mathbf{K}_+ = \mathbf{P}_+\mathbf{H}^T\,\mathbf{U}_+^{-1}.$$

Since, in this case, the updating steps (4b),(4d), and (4f) of the Kalman recursion (4), and the updating of the gain matrix (3), become redundant, the Kalman recursion (4) for $t = 1, 2, \dots, N$ reduces to

$$\mathbf{w}_{t|t-1,+} = \mathbf{F}\boldsymbol{\mu}_{t-1,+}, \tag{C.1a}$$
$$\mathbf{e}_{t,+} = \mathbf{y}_t^{(\mathbf{h})} - \mathbf{H}\mathbf{w}_{t|t-1,+}, \tag{C.1b}$$
$$\boldsymbol{\mu}_{t,+} = \mathbf{w}_{t|t-1,+} + \mathbf{K}_+\mathbf{e}_{t,+}. \tag{C.1c}$$

Defining

$$\mathbf{J}_+ := (\mathbf{I} - \mathbf{K}_+\mathbf{H})\,\mathbf{F},$$

[26]Note that from $\mathbf{C}_0 = \mathbf{F}\,\mathbf{C}_0\mathbf{F}^T + \mathbf{Q}$ and (4b) follows that $\mathbf{P}_{1|0} = \mathbf{F}\,\mathbf{C}_0\mathbf{F}^T + \mathbf{Q} = \mathbf{C}_0$.

we can use equations (C.1a) to (C.1c) to determine the law of motion for $\boldsymbol{\mu}_{t,+}$ as

$$\begin{aligned}
\boldsymbol{\mu}_{t,+} &= \mathbf{w}_{t|t-1,+} + \mathbf{K}_{+}\mathbf{e}_{t,+} \\
&= \mathbf{w}_{t|t-1,+} + \mathbf{K}_{+}\left(\mathbf{y}_t^{(\mathbf{h})} - \mathbf{H}\mathbf{w}_{t|t-1,+}\right) \\
&= \mathbf{F}\boldsymbol{\mu}_{t-1,+} + \mathbf{K}_{+}\left(\mathbf{y}_t^{(\mathbf{h})} - \mathbf{H}\mathbf{F}\boldsymbol{\mu}_{t-1,+}\right) \\
&= \mathbf{K}_{+}\,\mathbf{y}_t^{(\mathbf{h})} + (\mathbf{I}-\mathbf{K}_{+}\mathbf{H})\,\mathbf{F}\,\boldsymbol{\mu}_{t-1,+} \\
&= \mathbf{K}_{+}\,\mathbf{y}_t^{(\mathbf{h})} + \mathbf{J}_{+}\,\boldsymbol{\mu}_{t-1,+}, \qquad \forall t = 1,2,\ldots,N. \qquad \text{(C.2)}
\end{aligned}$$

This means we can use (C.2) to determine $\boldsymbol{\mu}_{t,+}$ for $t = 0,1,\ldots,N-1$ recursively. It follows directly from equations (C.1b) and (C.1c) that the quantities $\mathbf{w}_{t|t-1}$ and $\mathbf{e}_t$ for $t = 1,2,\ldots,N$ are then determined by

$$\begin{pmatrix}\mathbf{w}_{1|0,+} & \cdots & \mathbf{w}_{N|N-1,+}\end{pmatrix} = \mathbf{F}\begin{pmatrix}\boldsymbol{\mu}_{0,+} & \cdots & \boldsymbol{\mu}_{N-1,+}\end{pmatrix}, \qquad \text{(C.3)}$$

$$\begin{aligned}
\begin{pmatrix}\mathbf{e}_{1,+} & \cdots & \mathbf{e}_{N,+}\end{pmatrix} &= \begin{pmatrix}\mathbf{y}_1^{(\mathbf{h})} & \cdots & \mathbf{y}_N^{(\mathbf{h})}\end{pmatrix} - \mathbf{H}\begin{pmatrix}\mathbf{w}_{1|0,+} & \cdots & \mathbf{w}_{N|N-1,+}\end{pmatrix} \\
&= \begin{pmatrix}\mathbf{y}_1^{(\mathbf{h})} & \cdots & \mathbf{y}_N^{(\mathbf{h})}\end{pmatrix} - \mathbf{H}\mathbf{F}\begin{pmatrix}\boldsymbol{\mu}_{0,+} & \cdots & \boldsymbol{\mu}_{N-1,+}\end{pmatrix}. \qquad \text{(C.4)}
\end{aligned}$$

Note that we can also simplify the log-likelihood conditional to the initialization $(\boldsymbol{\mu}_{0,+}, \mathbf{C}_{+})$, since in this case $\mathbf{U}_t = \mathbf{U}_{+}$ for all $t = 1,\ldots,N$. Hence, we can define the log-likelihood computed based on the SKF as

$$\begin{aligned}
\log\left(f_{\mathbf{Y}_N}\right)_{+} &= -\frac{1}{2}\left(n_y\,N\,\log(2\pi) + \sum_{t=1}^{N}\log|\mathbf{U}_{+}| + \sum_{t=1}^{N}\mathbf{e}_{t,+}^T\,\mathbf{U}_{+}^{-1}\,\mathbf{e}_{t,+}\right) \\
&= -\frac{1}{2}\left(n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}|\right) - \frac{1}{2}\sum_{t=1}^{N}\operatorname{tr}\left(\mathbf{e}_{t,+}^T\,\mathbf{U}_{+}^{-1}\,\mathbf{e}_{t,+}\right) \\
&= -\frac{1}{2}\left(n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}|\right) - \frac{1}{2}\sum_{t=1}^{N}\operatorname{tr}\left(\mathbf{U}_{+}^{-1}\,\mathbf{e}_{t,+}\mathbf{e}_{t,+}^T\right) \\
&= -\frac{1}{2}\left(n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}|\right) - \frac{1}{2}\operatorname{tr}\left(\sum_{t=1}^{N}\mathbf{U}_{+}^{-1}\,\mathbf{e}_{t,+}\mathbf{e}_{t,+}^T\right) \\
&= -\frac{1}{2}\left(n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}|\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{U}_{+}^{-1}\,\sum_{t=1}^{N}\mathbf{e}_{t,+}\mathbf{e}_{t,+}^T\right) \\
&= -\frac{1}{2}\left[n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}| + \operatorname{tr}\left(\mathbf{U}_{+}^{-1}\begin{pmatrix}\mathbf{e}_{1,+} & \cdots & \mathbf{e}_{N,+}\end{pmatrix}\begin{pmatrix}\mathbf{e}_{1,+} & \cdots & \mathbf{e}_{N,+}\end{pmatrix}^T\right)\right] \\
&= -\frac{1}{2}\left[n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}| + \operatorname{tr}\left(\mathbf{U}_{+}^{-1}\,\mathbf{e}_{1:N,+}\mathbf{e}_{1:N,+}^T\right)\right], \\
&= -\frac{1}{2}\left[n_y\,N\,\log(2\pi) + N\,\log|\mathbf{U}_{+}| + \operatorname{tr}\left(\mathbf{e}_{1:N,+}^T\,\mathbf{U}_{+}^{-1}\,\mathbf{e}_{1:N,+}\right)\right], \qquad \text{(C.5)}
\end{aligned}$$

with $\mathbf{e}_{1:N,+} := \begin{pmatrix}\mathbf{e}_{1,+} & \cdots & \mathbf{e}_{N,+}\end{pmatrix}$.

Further, note that since $\mathbf{J}_{+} = \tilde{\mathbf{F}}_{\mathbf{C}_{+}}$, we may analyze its eigenvalues to check if $\mathbf{C}_{+}$ is a strong /

stabilizing solution to RDE (8a).

D Derivation of the augmented Kalman filter

The first part of this appendix contains the formal derivation of the augmented Kalman filter (AKF) (15) and the log-density $\log\left(f_{\mathbf{Y}_N}\right)$ given in (16). In the second part, we provide a brief digression on how initialization strategies for non-stationary SSMs, such as the *fixed-but-unknown* or the *diffuse* initialization, can be incorporated within the AKF (15). In the last part of this appendix, we show how to incorporate the additional steps of the AKF into the KF (4).

D.1 The augmented Kalman filter

Note that the derivation of the AKF given here in large parts follows the arguments of Durbin and Koopman (2012, Chapter 5.7). However, we shall derive the AKF with respect to SSM (1), while the elaborations of Durbin and Koopman (2012) are based on the alternative state-space representation (5).

For convenience, let us restate the model for the initial state vector from equation (13):

$$\mathbf{w}_0 = \mathbf{a}_w + \mathbf{A}_{\mathbf{w}}\,\bar{\mathbf{w}}_0 + \mathbf{A}_{\mathbf{d}}\,\mathbf{d}, \qquad \bar{\mathbf{w}}_0 \sim \mathrm{N}(\bar{\mu}_0, \bar{\mathbf{C}}_0), \qquad \mathbf{d} \sim \mathrm{N}(\boldsymbol{\delta}_0, \mathbf{D}_0),$$

Further, we denoted the time t quantities generated by the Kalman recursion (4) initialized at $(\tilde{\mu}_0, \tilde{\mathbf{C}}_0)$, with $\tilde{\mu}_0 = \mathbf{a}_w + \mathbf{A}_{\mathbf{w}}\bar{\mu}_0$ and $\tilde{\mathbf{C}}_0 = \mathbf{A}_{\mathbf{w}}\bar{\mathbf{C}}_0\mathbf{A}_{\mathbf{w}}^T$, by $\tilde{\mu}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$, and $\tilde{\mathbf{K}}_t$. In the following derivation of the AKF (15a)-(15e)

$$\begin{aligned}
\mu_t &= \tilde{\mu}_t + \mathbf{M}_t\mathbf{A}_d\,\left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_t\,\mathbf{A}_{\mathbf{d}}\right)^{-1}\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{s}_t\right), && \forall\ t = 1,2,\ldots,N,\\
\mathbf{C}_t &= \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_t\,\mathbf{A}_{\mathbf{d}}\right)^{-1}\mathbf{A}_d^T\mathbf{M}_t^T, && \forall\ t = 1,2,\ldots,N,\\
\mathbf{s}_t &= \mathbf{s}_{t-1} + (\mathbf{HFM}_{t-1})^T\,\tilde{\mathbf{U}}_t^{-1}\,\tilde{\mathbf{e}}_t, && \mathbf{s}_0 = \mathbf{0},\ \forall\ t = 1,2,\ldots,N,\\
\mathbf{S}_t &= \mathbf{S}_{t-1} + (\mathbf{HFM}_{t-1})^T\,\tilde{\mathbf{U}}_t^{-1}\,(\mathbf{HFM}_{t-1}), && \mathbf{S}_0 = \mathbf{0},\ \forall\ t = 1,2,\ldots,N,\\
\mathbf{M}_t &= \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{F}\,\mathbf{M}_{t-1}, && \mathbf{M}_0 = \mathbf{I},\ \forall\ t = 1,2,\ldots,N,
\end{aligned}$$

and especially of the log-density

$$\begin{aligned}
\log(f_{\mathbf{Y}_N}) = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) &- \frac{1}{2}\log|\mathbf{I} + \mathbf{D}_0\,\mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_N\,\mathbf{A}_{\mathbf{d}}| - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_{\mathbf{0}}^{-1}\boldsymbol{\delta}_0\\
&+ \frac{1}{2}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{s}_N)^T(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_N\,\mathbf{A}_{\mathbf{d}})^{-1}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\,\mathbf{s}_N),
\end{aligned}$$

given in (16), the Bayes theorem will play a key role as it allows us to decompose the log-density

$\log(f_{\mathbf{Y}_N})$ into

$$\log\left(f_{\mathbf{Y}_N}\right) = \log\left(\frac{f_{\mathbf{Y}_N|\mathbf{d}} \cdot f_{\mathbf{d}}}{f_{\mathbf{d}|\mathbf{Y}_N}}\right)$$
$$= \log(f_{\mathbf{d}}) + \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) - \log\left(f_{\mathbf{d}|\mathbf{Y}_N}\right). \tag{D.1}$$

While for given $\boldsymbol{\delta}_0$ and $\mathbf{D}_0$, the log-density $\log(f_{\mathbf{d}})$ is fully specified, we need to obtain $\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)$ and $\log\left(f_{\mathbf{d}|\mathbf{Y}_N}\right)$ to determine $\log\left(f_{\mathbf{Y}_N}\right)$ from the right-hand side of (D.1). Therefore, in the first step, we will show that we can express the log-density $\log(f_{\mathbf{Y}_N|\mathbf{d}})$ as a function of $\log(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}})$, $\mathbf{d}$, $\mathbf{A}_d$, $\mathbf{s}_N$, and $\mathbf{S}_N$, where a crucial preliminary result will be the observation that the quantities $\boldsymbol{\mu}_t$, $\mathbf{w}_{t|t-1}$, and $\mathbf{e}_t$ of the KF are linear functions of $\boldsymbol{\mu}_0$. This observation is due to Rosenberg (1973) and forms the basis for the so-called fixed-but-unknown initialization discussed later in this appendix. Eventually, we will receive (15) and (16) using a fixed-point smoothing algorithm to obtain $\log\left(f_{\mathbf{d}|\mathbf{Y}_N}\right)$.

Linearity of $\boldsymbol{\mu}_t$, $\mathbf{w}_{t|t-1}$, and $\mathbf{e}_t$ in $\boldsymbol{\mu}_0$: In the following lemma, we will show that the quantities $\boldsymbol{\mu}_t$, $\mathbf{w}_{t|t-1}$, and $\mathbf{e}_t$ of the KF are linear in $\boldsymbol{\mu}_0$, while the quantities $\mathbf{C}_t$, $\mathbf{P}_{t|t-1}$, and $\mathbf{U}_t$ are independent of $\boldsymbol{\mu}_0$:

Lemma D.1 *Suppose for the SSM (1) the time t quantities generated by the Kalman recursion (4) initialized at* $(\tilde{\boldsymbol{\mu}}_0, \mathbf{C}_0)$ *are denoted by* $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$ *and* $\tilde{\mathbf{K}}_t$. *Further suppose* $\boldsymbol{\mu}_t$, $\mathbf{C}_t$, $\mathbf{w}_{t|t-1}$, $\mathbf{P}_{t|t-1}$, $\mathbf{e}_t$, $\mathbf{U}_t$ *and* $\mathbf{K}_t$ *denote the time t quantities generated by the Kalman recursion (4) initialized at* $(\boldsymbol{\mu}_0, \mathbf{C}_0)$, *then we can state that*

$$\mathbf{C}_t = \tilde{\mathbf{C}}_t, \qquad \mathbf{P}_{t|t-1} = \tilde{\mathbf{P}}_{t|t-1}, \qquad \mathbf{U}_t = \tilde{\mathbf{U}}_t, \qquad \mathbf{K}_t = \tilde{\mathbf{K}}_t,$$

and that

$$\boldsymbol{\mu}_t = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\,\boldsymbol{\Delta}_0, \qquad \mathbf{w}_{t|t-1} = \tilde{\mathbf{w}}_{t|t-1} + \mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0, \qquad \mathbf{e}_t = \tilde{\mathbf{e}}_t - \mathbf{H}\mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0,$$

with

$$\boldsymbol{\Delta}_0 = \boldsymbol{\mu}_0 - \tilde{\boldsymbol{\mu}}_0, \qquad \mathbf{M}_s = \prod_{j=1}^{s} \mathbf{J}_{s-j+1}, \qquad \mathbf{J}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\,\mathbf{F}, \qquad \forall s = 0, 1, \ldots, N,$$

for all $t = 1, 2, \ldots, N$.

Proof:

Defining $\mathbf{J}_t := (\mathbf{I} - \mathbf{K}_t\mathbf{H})\,\mathbf{F}$ we can use equation (4a), (4c) and (4e) to obtain the law of motion for $\boldsymbol{\mu}_t$ as

$$\begin{aligned}
\boldsymbol{\mu}_t &= \mathbf{w}_{t|t-1} + \mathbf{K}_t \mathbf{e}_t \\
&= \mathbf{w}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{h} - \mathbf{H}\, \mathbf{w}_{t|t-1}\right) \\
&= \mathbf{w}_{t|t-1} + \mathbf{K}_t \left(\mathbf{y}_t - \mathbf{h}\right) - \mathbf{K}_t \mathbf{H}\, \mathbf{w}_{t|t-1} \\
&= \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}\right) \mathbf{w}_{t|t-1} + \mathbf{K}_t\ \left(\mathbf{y}_t - \mathbf{h}\right) \\
&= \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}\right) \mathbf{F}\, \boldsymbol{\mu}_{t-1} + \mathbf{K}_t\ \left(\mathbf{y}_t - \mathbf{h}\right) \\
&= \mathbf{J}_t\ \boldsymbol{\mu}_{t-1} + \mathbf{K}_t\ \left(\mathbf{y}_t - \mathbf{h}\right), \qquad \forall t = 1, 2, \ldots, N.
\end{aligned} \tag{D.2}$$

Note that it follows from equations (3), (4b), (4d) and (4f), that the sequences $\{\mathbf{K}_t\}_{t=1}^N$, $\{\mathbf{P}_{t|t-1}\}_{t=1}^N$, $\{\mathbf{U}_t\}_{t=1}^N$ and $\{\mathbf{C}_t\}_{t=0}^N$ referring to the initialization $(\boldsymbol{\mu}_0, \mathbf{C}_0)$ do not depend on $\boldsymbol{\mu}_0$ and therefore are identical to the sequences $\{\tilde{\mathbf{K}}_t\}_{t=1}^N$, $\{\tilde{\mathbf{P}}_{t|t-1}\}_{t=1}^N$, $\{\tilde{\mathbf{U}}_t\}_{t=1}^N$ and $\{\tilde{\mathbf{C}}_t\}_{t=0}^N$ referring to the initialization $(\tilde{\boldsymbol{\mu}}_0, \mathbf{C}_0)$. Consequently, we can state that law of motion for $\tilde{\boldsymbol{\mu}}_t$ similar to (D.2) is given by

$$\tilde{\boldsymbol{\mu}}_t = \mathbf{J}_t\ \tilde{\boldsymbol{\mu}}_{t-1} + \mathbf{K}_t\ \left(\mathbf{y}_t - \mathbf{h}\right), \qquad \forall t = 1, 2, \ldots, N. \tag{D.3}$$

Moreover, defining $\boldsymbol{\Delta}_t = \boldsymbol{\mu}_t - \tilde{\boldsymbol{\mu}}_t$ for all $t = 0, 1, \ldots, N$, we may use (D.2) and (D.3) to write

$$\begin{aligned}
\boldsymbol{\Delta}_t &= \mathbf{J}_t\ \boldsymbol{\Delta}_{t-1} \\
&= \mathbf{J}_t \cdot \mathbf{J}_{t-1}\ \boldsymbol{\Delta}_{t-2} \\
&\ \vdots \\
&= \mathbf{J}_t \cdot \mathbf{J}_{t-1} \cdot \ldots \cdot \mathbf{J}_1\ \boldsymbol{\Delta}_0 \\
&= \left(\prod_{j=1}^{t} \mathbf{J}_{t-j+1}\right) \boldsymbol{\Delta}_0 \\
&= \mathbf{M}_t\ \boldsymbol{\Delta}_0, \qquad \forall t = 0, 1, \ldots, N,
\end{aligned} \tag{D.4}$$

with

$$\mathbf{M}_t := \prod_{j=1}^{t} \mathbf{J}_{t-j+1}, \qquad \forall t = 0, 1, \ldots, N.$$

Note that from the definition of the $\prod(\cdot)$ operator follows that $\mathbf{M}_0 = \prod_{j=1}^{0} \mathbf{J}_{j+1} = \mathbf{I}$. The statement of Lemma D.1 then follows directly from (D.4), (4a) and (4b):

$$\boldsymbol{\mu}_t = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\ \boldsymbol{\Delta}_0, \qquad \forall t = 0, 1, \ldots, N, \tag{D.5}$$

$$\mathbf{w}_{t|t-1} = \mathbf{F}\ \boldsymbol{\mu}_{t-1}$$

$$
\begin{aligned}
&= \mathbf{F}\,\tilde{\boldsymbol{\mu}}_{t-1} + \mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0 \\
&= \tilde{\mathbf{w}}_{t|t-1} + \mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0, \qquad \forall t = 1,2,\ldots,N, \qquad \text{(D.6)}\\
\mathbf{e}_t &= \mathbf{y}_t - \mathbf{h} - \mathbf{H}\mathbf{w}_{t|t-1} \\
&= \mathbf{y}_t - \mathbf{h} - \mathbf{H}\left(\tilde{\mathbf{w}}_{t|t-1} + \mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0\right) \\
&= \mathbf{y}_t - \mathbf{h} - \mathbf{H}\tilde{\mathbf{w}}_{t|t-1} - \mathbf{H}\mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0 \\
&= \tilde{\mathbf{e}}_t - \mathbf{H}\mathbf{F}\mathbf{M}_{t-1}\,\boldsymbol{\Delta}_0, \qquad \forall t = 1,2,\ldots,N. \qquad \text{(D.7)}
\end{aligned}
$$

□

As mentioned before, this observation, closely connected to the so-called fixed-but-unknown initialization, where we treat the elements of $\mathbf{d}$ as fixed parameters that we may estimate via maximum-likelihood, is due to Rosenberg (1973).

Obtaining $\log(f_{\mathbf{Y}_N|\mathbf{d}})$ **as a function of** $\log(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}})$**,** $\mathbf{d}$**,** $\mathbf{A}_d$**,** $\mathbf{s}_N$**, and** $\mathbf{S}_N$**:** In the following Proposition, we use Lemma D.1 to obtain the analytical maximum-likelihood estimator for $\mathbf{d}$ and show that we may rewrite $\log(f_{\mathbf{Y}_N|\mathbf{d}})$ as a function of $\log(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}})$, $\mathbf{d}$, $\mathbf{A}_d$, $\mathbf{s}_N$, and $\mathbf{S}_N$:

Proposition D.1 *Suppose the initial state vector* $\mathbf{w}_0$ *can be written as*

$$
\mathbf{w}_0 = \mathbf{a}_w + \mathbf{A}_{\mathbf{w}}\,\bar{\mathbf{w}}_0 + \mathbf{A}_{\mathbf{d}}\,\mathbf{d}, \qquad \bar{\mathbf{w}}_0 \sim N(\bar{\boldsymbol{\mu}}_0, \bar{\mathbf{C}}_0), \qquad \mathbf{d} \sim N(\boldsymbol{\delta}_0, \mathbf{D}_0),
$$

where $\bar{\mathbf{w}}_0 \in \mathbb{R}^{n_{\bar{w}}}$, $n_{\bar{w}} \leq n_w$ *and* $\mathbf{d} \in \mathbb{R}^{n_d}$, $n_d \leq n_w$ *represent two independent random vectors. Suppose for SSM* (1) *the time* t *quantities generated by the Kalman recursion* (4) *initialized at* $(\tilde{\boldsymbol{\mu}}_0, \tilde{\mathbf{C}}_0)$*, with* $\tilde{\boldsymbol{\mu}}_0 = \mathbf{a}_w + \mathbf{A}_{\mathbf{w}}\bar{\boldsymbol{\mu}}_0$ *and* $\tilde{\mathbf{C}}_0 = \mathbf{A}_{\mathbf{w}}\bar{\mathbf{C}}_0\mathbf{A}_{\mathbf{w}}^T$*, are denoted by* $\tilde{\boldsymbol{\mu}}_t$*,* $\tilde{\mathbf{C}}_t$*,* $\tilde{\mathbf{w}}_{t|t-1}$*,* $\tilde{\mathbf{P}}_{t|t-1}$*,* $\tilde{\mathbf{e}}_t$*,* $\tilde{\mathbf{U}}_t$ *and* $\tilde{\mathbf{K}}_t$*. Then for the SSM* (1) *the conditional log-density of* $\mathbf{Y}_N$ *given* $\mathbf{d}$ *may be written as*

$$
\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \mathbf{d}^T\,\mathbf{A}_{\mathbf{d}}^T\,\mathbf{s}_N - \frac{1}{2}\,\mathbf{d}^T\,\mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_N\,\mathbf{A}_{\mathbf{d}}\,\mathbf{d},
$$

with

$$
\begin{aligned}
&\mathbf{s}_t = \sum_{i=1}^{N} \mathbf{E}_i^T\tilde{\mathbf{U}}_i^{-1}\,\tilde{\mathbf{e}}_i, \qquad \mathbf{S}_t = \sum_{i=1}^{N} \mathbf{E}_i^T\tilde{\mathbf{U}}_i^{-1}\mathbf{E}_i \\
&\mathbf{E}_t = \mathbf{H}\mathbf{F}\mathbf{M}_{t-1}, \qquad \mathbf{M}_{t-1} = \prod_{j=1}^{t-1}\mathbf{J}_{t-j}, \qquad \mathbf{J}_t = \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{F}, \qquad \forall t = 1,2,\ldots,N.
\end{aligned}
$$

Further, the maximum-likelihood estimator of $\mathbf{d}$ *for a given sample* $\mathbf{Y}_N$ *yields*

$$
\hat{\mathbf{d}} = \underset{\mathbf{d}}{\operatorname{argmax}} \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = \left(\mathbf{A}_{\mathbf{d}}^T\,\mathbf{S}_N\,\mathbf{A}_{\mathbf{d}}\right)^{-1}\,\mathbf{A}_{\mathbf{d}}^T\,\mathbf{s}_N,
$$

with

$$\widehat{\mathrm{Var}}[\hat{\mathbf{d}}] = -\left(\frac{\partial^2 \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)}{\partial \mathbf{d}\, \partial \mathbf{d}^T}\right)^{-1} = \left(\mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_N\ \mathbf{A}_{\mathbf{d}}\right)^{-1}.$$

Proof:

Note that we may write the mean vector and the variance matrix of $\mathbf{w}_0$ given $\mathbf{d}$ as

$$\boldsymbol{\mu}_{0|\mathbf{d}} := \mathrm{E}[\mathbf{w}_0|\mathbf{d}] = \tilde{\boldsymbol{\mu}}_0 + \bar{\mathbf{d}}, \qquad \mathbf{C}_{0|\mathbf{d}} := \mathrm{Var}[\mathbf{w}_0|\mathbf{d}] = \tilde{\mathbf{C}}_0,$$

with $\bar{\mathbf{d}} = \mathbf{A}_{\mathbf{d}}\ \mathbf{d}$. Suppose for SSM (1) the time t quantities generated by the Kalman recursion (4) initialized at $(\boldsymbol{\mu}_{0|\mathbf{d}}, \mathbf{C}_{0|\mathbf{d}})$ are denoted by $\boldsymbol{\mu}_{t|\mathbf{d}}$, $\mathbf{C}_{t|\mathbf{d}}$, $\mathbf{w}_{t|t-1,\mathbf{d}}$, $\mathbf{P}_{t|t-1,\mathbf{d}}$, $\mathbf{e}_{t|\mathbf{d}}$, $\mathbf{U}_{t|\mathbf{d}}$ and $\mathbf{K}_{t|\mathbf{d}}$. Then the conditional log-density of $\mathbf{Y}_N$ given $\mathbf{d}$ follows from (6) as

$$\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = -\frac{N\ n_y}{2}\ \log(2\pi)\ - \frac{1}{2}\sum_{t=1}^{N} \log\left|\mathbf{U}_{t|\mathbf{d}}\right|\ - \frac{1}{2}\sum_{t=1}^{N} \mathbf{e}_{t|\mathbf{d}}^T\ \mathbf{U}_{t|\mathbf{d}}^{-1}\ \mathbf{e}_{t|\mathbf{d}}.$$

Hence, using Lemma D.1 we may write

$$\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = -\frac{N\ n_y}{2}\ \log(2\pi)\ - \frac{1}{2}\sum_{t=1}^{N} \log\left|\tilde{\mathbf{U}}_t\right| - \frac{1}{2}\sum_{t=1}^{N}\left(\tilde{\mathbf{e}}_t - \mathbf{E}_t\ \bar{\mathbf{d}}\right)^T \tilde{\mathbf{U}}_t^{-1}\left(\tilde{\mathbf{e}}_t - \mathbf{E}_t\ \bar{\mathbf{d}}\right). \tag{D.8}$$

The first part of the proof is completed by rewriting the (D.8) to

$$\begin{aligned}
\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) &= -\frac{N\ n_y}{2}\ \log(2\pi)\ - \frac{1}{2}\sum_{t=1}^{N} \log\left|\tilde{\mathbf{U}}_t\right| - \frac{1}{2}\sum_{t=1}^{N}\left(\tilde{\mathbf{e}}_t - \mathbf{E}_t\ \bar{\mathbf{d}}\right)^T \tilde{\mathbf{U}}_t^{-1}\left(\tilde{\mathbf{e}}_t - \mathbf{E}_t\ \bar{\mathbf{d}}\right)\\
&= \underbrace{-\frac{N\ n_y}{2}\ \log(2\pi)\ - \frac{1}{2}\sum_{t=1}^{N} \log\left|\tilde{\mathbf{U}}_t\right|\ - \frac{1}{2}\left(\sum_{t=1}^{N} \tilde{\mathbf{e}}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \tilde{\mathbf{e}}_t\right)}_{=\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)}\\
&\quad + \frac{1}{2}\left(\sum_{t=1}^{N} \tilde{\mathbf{e}}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \mathbf{E}_t\ \bar{\mathbf{d}}\right) + \frac{1}{2}\left(\sum_{t=1}^{N} \bar{\mathbf{d}}^T\ \mathbf{E}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \tilde{\mathbf{e}}_t\right) - \frac{1}{2}\left(\sum_{t=1}^{N} \bar{\mathbf{d}}^T\ \mathbf{E}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \mathbf{E}_t\ \bar{\mathbf{d}}\right)\\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)\\
&\quad + \frac{1}{2}\left(\sum_{t=1}^{N} \tilde{\mathbf{e}}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \mathbf{E}_t\right)\bar{\mathbf{d}} + \frac{1}{2}\bar{\mathbf{d}}^T\left(\sum_{t=1}^{N} \mathbf{E}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \tilde{\mathbf{e}}_t\right) - \frac{1}{2}\bar{\mathbf{d}}^T\left(\sum_{t=1}^{N} \mathbf{E}_t^T\ \tilde{\mathbf{U}}_t^{-1}\ \mathbf{E}_t\right)\bar{\mathbf{d}}\\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)\ \underbrace{+\frac{1}{2}\mathbf{s}_N^T\bar{\mathbf{d}} + \frac{1}{2}\bar{\mathbf{d}}^T\ \mathbf{s}_N}_{=\bar{\mathbf{d}}^T\mathbf{s}_N,\ \text{since}\ \left(\bar{\mathbf{d}}^T\mathbf{s}_N\right)\in\mathbb{R}^{1\times 1}.}\ -\frac{1}{2}\bar{\mathbf{d}}^T\ \mathbf{S}_N\ \bar{\mathbf{d}}\\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)\ + \bar{\mathbf{d}}^T\mathbf{s}_N - \frac{1}{2}\bar{\mathbf{d}}^T\ \mathbf{S}_N\ \bar{\mathbf{d}}\\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)\ + \mathbf{d}^T\mathbf{A}_{\mathbf{d}}^T\ \mathbf{s}_N - \frac{1}{2}\ \mathbf{d}^T\ \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_N\ \mathbf{A}\ \mathbf{d}.
\end{aligned} \tag{D.9}$$

Since $\mathbf{s}_N$, $\mathbf{S}_N$, $\mathbf{A_d}$ and $\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right)$ do not depend on $\mathbf{d}$, using matrix differentiating rules (see e.g. Lütkepohl (2007, pp. 664-671)) the first and second order derivatives of $\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)$ with respect to $\mathbf{d}$ yield

$$\frac{\partial \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)}{\partial \mathbf{d}} = \mathbf{A_d}^T\ \mathbf{s}_N - \mathbf{A_d}^T\ \mathbf{S}_N\ \mathbf{A_d}\ \mathbf{d}, \qquad \text{(D.10a)} \qquad \frac{\partial \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)}{\partial \mathbf{d}\, \partial \mathbf{d}^T} = -\mathbf{A_d}^T\ \mathbf{S}_N\ \mathbf{A_d}. \qquad \text{(D.10b)}$$

If the matrix $\mathbf{A_d}^T\ \mathbf{S}_N\ \mathbf{A_d}$ has full rank, equating (D.10a) to zero yields the maximum-likelihood estimator for $\mathbf{d}$ given $(\mathbf{Y}_N)$

$$\hat{\mathbf{d}} = \underset{\mathbf{d}}{\operatorname{argmax}} \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = \left(\mathbf{A_d}^T\ \mathbf{S}_N\ \mathbf{A_d}\right)^{-1}\ \mathbf{A_d}^T\ \mathbf{s}_N,$$

with

$$\widehat{\operatorname{Var}}[\hat{\mathbf{d}}] = -\left(\frac{\partial^2 \log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right)}{\partial \mathbf{d}\, \partial \mathbf{d}^T}\right)^{-1} = \left(\mathbf{A_d}^T\ \mathbf{S}_N\ \mathbf{A_d}\right)^{-1}.$$

□

Derivation of (15) and (16): Using the results of Lemma D.1 and Proposition D.1, we may ultimately obtain the formulas of the AKF (15) and the log-density $\log(f_{\mathbf{Y}_N})$ given in (16). Therefore, let us establish the following Proposition:

Proposition D.2 *Suppose the initial state vector $\mathbf{w}_0$ can be written as*

$$\mathbf{w}_0 = \mathbf{a}_w + \mathbf{A_w}\ \bar{\mathbf{w}}_0 + \mathbf{A_d}\ \mathbf{d}, \qquad \bar{\mathbf{w}}_0 \sim N(\bar{\boldsymbol{\mu}}_0, \bar{\mathbf{C}}_0), \qquad \mathbf{d} \sim N(\boldsymbol{\delta}_0, \mathbf{D}_0),$$

where $\bar{\mathbf{w}}_0 \in \mathbb{R}^{n_{\bar{w}}}$, $n_{\bar{w}} \le n_w$ and $\mathbf{d} \in \mathbb{R}^{n_d}$, $n_d \le n_w$ represent two independent random vectors. Suppose for SSM (1) the time t quantities generated by the Kalman recursion (4) initialized at $(\tilde{\boldsymbol{\mu}}_0, \tilde{\mathbf{C}}_0)$, with $\tilde{\boldsymbol{\mu}}_0 = \mathbf{a}_w + \mathbf{A_w}\bar{\boldsymbol{\mu}}_0$ and $\tilde{\mathbf{C}}_0 = \mathbf{A_w}\bar{\mathbf{C}}_0\mathbf{A_w}^T$, are denoted by $\tilde{\boldsymbol{\mu}}_t$, $\tilde{\mathbf{C}}_t$, $\tilde{\mathbf{w}}_{t|t-1}$, $\tilde{\mathbf{P}}_{t|t-1}$, $\tilde{\mathbf{e}}_t$, $\tilde{\mathbf{U}}_t$ and $\tilde{\mathbf{K}}_t$. Suppose the variance matrix $\mathbf{D}_0$ is positive definite. Then for the SSM (1) the conditional distributions of $\mathbf{d}$ and $\mathbf{w}_t$ given $\mathbf{Y}_t$ are Gaussian with mean vectors

$$\begin{aligned}
\boldsymbol{\delta}_t &:= \mathrm{E}[\mathbf{d}|\mathbf{Y}_t] = \left(\mathbf{D}_0^{-1} + \mathbf{A_d}^T\ \mathbf{S}_t\ \mathbf{A_d}\right)^{-1}\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A_d}^T\ \mathbf{s}_t\right),\\
\boldsymbol{\mu}_t &:= \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_t] = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\mathbf{A}_d\ \boldsymbol{\delta}_t,
\end{aligned}$$

and variance matrices:

$$\begin{aligned}
\mathbf{D}_t &:= \operatorname{Var}[\mathbf{d}|\mathbf{Y}_t] = \left(\mathbf{D}_0^{-1} + \mathbf{A_d}^T\ \mathbf{S}_t\ \mathbf{A_d}\right)^{-1},\\
\mathbf{C}_t &:= \operatorname{Var}[\mathbf{w}_t|\mathbf{Y}_t] = \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T,
\end{aligned}$$

where

$$\mathbf{s}_t = \sum_{i=1}^{t} \mathbf{E}_i^T \tilde{\mathbf{U}}_i^{-1} \tilde{\mathbf{e}}_i, \qquad \mathbf{S}_t = \sum_{i=1}^{t} \mathbf{E}_i^T \tilde{\mathbf{U}}_i^{-1} \mathbf{E}_i \qquad \textit{and}$$

$$\mathbf{E}_t = \mathbf{HFM}_{t-1}, \qquad \mathbf{M}_{t-1} = \prod_{j=1}^{t-1} \mathbf{J}_{t-j}, \quad \mathbf{J}_t = \left(\mathbf{I} - \tilde{\mathbf{K}}_t \mathbf{H}\right)\mathbf{F}, \quad \forall t = 1, 2, \ldots, N.$$

Further, the log-density of $\mathbf{Y}_N$ *may be written as*

$$\log(f_{\mathbf{Y}_N}) = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\log|\mathbf{I} + \mathbf{D}_0\, \mathbf{A}_{\mathbf{d}}^T\, \mathbf{S}_N\, \mathbf{A}_{\mathbf{d}}| - \frac{1}{2}\boldsymbol{\delta}_0^T \mathbf{D}_{\mathbf{0}}^{-1} \boldsymbol{\delta}_0$$
$$+ \frac{1}{2}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\, \mathbf{s}_N)^T (\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\, \mathbf{S}_N\, \mathbf{A}_{\mathbf{d}})^{-1} (\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\, \mathbf{s}_N).$$

Proof:
From Lemma A.1 and the linearity of the SSM (1) follows that $\mathbf{d}$ given $\mathbf{Y}_t$ is normally distributed with the corresponding log-density

$$\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right) = -\frac{n_d}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{D}_t| - \frac{1}{2}(\mathbf{d} - \boldsymbol{\delta}_t)^T\, \mathbf{D}_t^{-1}\, (\mathbf{d} - \boldsymbol{\delta}_t), \tag{D.11}$$

where $\boldsymbol{\delta}_t := \mathrm{E}[\mathbf{d}|\mathbf{Y}_t]$ and $\mathbf{D}_t := \mathrm{Var}[\mathbf{d}|\mathbf{Y}_t]$ denote mean vector and the variance matrix of $\mathbf{d}$ given $\mathbf{Y}_t$. Since $\mathbf{d}$ given $\mathbf{Y}_t$ is normally distributed, the mode and the mean of $\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)$ coincide and we can write

$$\boldsymbol{\delta}_t = \underset{\mathbf{d}}{\operatorname{argmax}}\ \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right).$$

Additionally from (D.11) and matrix differentiating rules (see e.g. Lütkepohl (2007, pp. 664-671)) follows that the Hessian matrix of $\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)$ with respect to $\mathbf{d}$ is

$$\begin{aligned}
\frac{\partial^2 \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)}{\partial \mathbf{d}\, \partial \mathbf{d}^T} &= \frac{\partial}{\partial \mathbf{d}}\left(-\frac{1}{2}\frac{\partial\, (\mathbf{d} - \boldsymbol{\delta}_t)^T\, \mathbf{D}_t^{-1}\, (\mathbf{d} - \boldsymbol{\delta}_t)}{\partial \mathbf{d}^T}\right) \\
&= \frac{\partial}{\partial \mathbf{d}}\left(-\frac{1}{2}\frac{\partial\, (\boldsymbol{\delta}_t - \mathbf{d})^T\, \mathbf{D}_t^{-1}\, (\boldsymbol{\delta}_t - \mathbf{d})}{\partial \mathbf{d}^T}\right) \\
&= \frac{\partial}{\partial \mathbf{d}}\left(-\frac{1}{2}\left(-2\,(\boldsymbol{\delta}_t - \mathbf{d})^T\, \mathbf{D}_t^{-1}\right)\right) \\
&= \frac{\partial}{\partial \mathbf{d}}\left((\boldsymbol{\delta}_t - \mathbf{d})^T\, \mathbf{D}_t^{-1}\right) \\
&= -\mathbf{D}_t^{-1},
\end{aligned}$$

and the variance matrix $\mathbf{D}_t$ yields

$$\mathbf{D}_t = -\left(\frac{\partial^2 \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)}{\partial \mathbf{d}\, \partial \mathbf{d}^T} \right)^{-1}. \tag{D.12}$$

Hence, to prove the first part of Proposition D.2 we need to obtain the first and second order derivatives of $\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)$ with respect to $\mathbf{d}$. To do so, we first use the Bayes Theorem to rewrite $\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)$ as

$$\begin{aligned}\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right) &= \log\left(\frac{f_{\mathbf{Y}_t|\mathbf{d}} \cdot f_{\mathbf{d}}}{f_{\mathbf{Y}_t}}\right) \\ &= \log(f_{\mathbf{d}}) + \log\left(f_{\mathbf{Y}_t|\mathbf{d}}\right) - \log\left(f_{\mathbf{Y}_t}\right). \end{aligned} \tag{D.13}$$

Note that the first term on the right-hand side of (D.13) is the log-density of $\mathbf{d}$, which yields

$$\log(f_{\mathbf{d}}) = -\frac{n_d}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{D}_0| - \frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_0)^T\ \mathbf{D}_0^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_0). \tag{D.14}$$

Differentiating equation (D.14) with respect to $\mathbf{d}$ we get

$$\begin{aligned}\frac{\partial \log(f_{\mathbf{d}})}{\partial \mathbf{d}} &= -\frac{1}{2}\frac{\partial}{\partial \mathbf{d}}\left((\mathbf{d}-\boldsymbol{\delta}_0)^T\ \mathbf{D}_0^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_0)\right) \\ &= -\frac{1}{2}\frac{\partial}{\partial \mathbf{d}}\left((\boldsymbol{\delta}_0-\mathbf{d})^T\ \mathbf{D}_0^{-1}\ (\boldsymbol{\delta}_0-\mathbf{d})\right) \\ &= -\frac{1}{2}\left(-2\mathbf{D}_0^{-1}(\boldsymbol{\delta}_0-\mathbf{d})\right) \\ &= \mathbf{D}_0^{-1}(\boldsymbol{\delta}_0-\mathbf{d}) \\ &= \mathbf{D}_0^{-1}\boldsymbol{\delta}_0 - \mathbf{D}_0^{-1}\mathbf{d}. \end{aligned} \tag{D.15}$$

Furthermore, we already obtained the first order derivatives of the conditional log-likelihood $\log\left(f_{\mathbf{Y}_t|\mathbf{d}}\right)$ with respect to $\mathbf{d}$ in equation (D.10a). Finally, we get the first order derivatives of $\log\left(f_{\mathbf{Y}_t}\right)$ with respect to $\mathbf{d}$ as

$$\frac{\partial \log\left(f_{\mathbf{Y}_t}\right)}{\partial \mathbf{d}} = 0, \tag{D.16}$$

since we can obtain $\log\left(f_{\mathbf{Y}_t}\right)$ without knowledge of $\mathbf{d}$ from equation (6) by initializing the Kalman recursion (4) at $\left(\boldsymbol{\mu}_0, \mathbf{C}_0\right)$ with $\boldsymbol{\mu}_0 = \mathbf{a}_w + \mathbf{A}_w\ \bar{\boldsymbol{\mu}}_0 + \mathbf{A}_d\ \boldsymbol{\delta}_0$ and $\mathbf{C}_0 = \mathbf{A}_w\ \bar{\mathbf{C}}_0\ \mathbf{A}_w^T + \mathbf{A}_d\ \mathbf{D}_0\ \mathbf{A}_d^T$.

Hence, from equations (D.10a), (D.13), (D.15) and (D.16) we receive

$$\begin{aligned}\frac{\partial \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)}{\partial \mathbf{d}} &= \frac{\partial \log\left(f_{\mathbf{d}}\right)}{\partial \mathbf{d}} + \frac{\partial \log\left(f_{\mathbf{Y}_t|\mathbf{d}}\right)}{\partial \mathbf{d}} - \frac{\partial \log\left(f_{\mathbf{Y}_t}\right)}{\partial \mathbf{d}} \\ &= \mathbf{D}_0^{-1}\boldsymbol{\delta}_0 - \mathbf{D}_0^{-1}\mathbf{d} + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{s}_t - \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_t\ \mathbf{A}_{\mathbf{d}}\ \mathbf{d} \\ &= \mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{s}_t - \left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_t\ \mathbf{A}_{\mathbf{d}}\right)\mathbf{d},\end{aligned} \tag{D.17}$$

and the mean vector $\boldsymbol{\delta}_t$ of $\mathbf{d}$ given $\mathbf{Y}_t$ is obtained as

$$\begin{aligned}\boldsymbol{\delta}_t &= \underset{\mathbf{d}}{\operatorname{argmax}}\ \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right) \\ &= \left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_t\ \mathbf{A}_{\mathbf{d}}\right)^{-\mathbf{1}}\ \left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{s}_t\right),\end{aligned}$$

by equating (D.17) to zero and solving with respect to $\mathbf{d}$. To obtain $\mathbf{D}_t$, we compute the Hessian of $\log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)$ with respect to $\mathbf{d}$ by differentiating (D.17) with respect to $\mathbf{d}^T$, which results in

$$\frac{\partial \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)}{\partial \mathbf{d}\ \partial \mathbf{d}^T} = -\left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_t\ \mathbf{A}_{\mathbf{d}}\right). \tag{D.18}$$

Thus, due to (D.12), the variance matrix $\mathbf{D}_t$ of $\mathbf{d}$ given $\mathbf{Y}_t$ equals

$$\mathbf{D}_t = -\left(\frac{\partial^2 \log\left(f_{\mathbf{d}|\mathbf{Y}_t}\right)}{\partial \mathbf{d}\ \partial \mathbf{d}^T}\right)^{-1} = \left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T\ \mathbf{S}_t\ \mathbf{A}_{\mathbf{d}}\right)^{-1}.$$

Moreover, we can state that $\begin{pmatrix}\mathbf{d}\\ \mathbf{w}_t\end{pmatrix}$ given $\mathbf{Y}_t$ is normally distributed with mean vector

$$\mathrm{E}\left(\begin{matrix}\mathbf{d}\\ \mathbf{w}_t\end{matrix}\ \middle|\mathbf{Y}_t\right) = \begin{pmatrix}\boldsymbol{\delta}_t\\ \tilde{\mu}_t + \mathbf{M}_t\mathbf{A}_d\ \boldsymbol{\delta}_t\end{pmatrix} \tag{D.19}$$

and variance matrix

$$\mathrm{Var}\left(\begin{matrix}\mathbf{d}\\ \mathbf{w}_t\end{matrix}\ \middle|\mathbf{Y}_t\right) = \begin{pmatrix}\mathbf{D}_t & \mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T\\ \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t & \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T\end{pmatrix}. \tag{D.20}$$

This may be seen from the fact that we can write the joint density function of $\mathbf{d}$ and $\mathbf{w}_t$ given $\mathbf{Y}_t$ as

$$\begin{aligned}f_{\mathbf{d},\mathbf{w}_t|\mathbf{Y}_t} =& f_{\mathbf{d}|\mathbf{Y}_t}\cdot f_{\mathbf{w}_t|\mathbf{d},\mathbf{Y}_t} \\ =& (2\pi)^{-\frac{n_p}{2}}\ |\mathbf{D}_t|^{-\frac{1}{2}}\ \exp\left(-\frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_t)^T\ \mathbf{D}_t^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_t)\right)\end{aligned}$$

$$
\begin{aligned}
&\times (2\pi)^{-\frac{n_w}{2}} \left|\tilde{\mathbf{C}}_t\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d}\right)^T \tilde{\mathbf{C}}_t^{-1} \left(\mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d}\right)\right) \\
=&(2\pi)^{-\frac{n_p+n_w}{2}} \left|\begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t \end{pmatrix}\right|^{-\frac{1}{2}} \\
&\times \exp\left[-\frac{1}{2}\begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}^T \begin{pmatrix} \mathbf{D}_t^{-1} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}\right] \\
=&(2\pi)^{-\frac{n_p+n_w}{2}} \left(\underbrace{\left|\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_t\mathbf{A}_d & \mathbf{I} \end{pmatrix}\right|}_{=1} \left|\begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t \end{pmatrix}\right| \underbrace{\left|\begin{pmatrix} \mathbf{I} & \mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right|}_{=1}\right)^{-\frac{1}{2}} \\
&\times \exp\Bigg[-\frac{1}{2}\begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}^T \underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}^{-1}}_{=\mathbf{I}} \\
&\qquad \times \begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t \end{pmatrix}^{-1} \underbrace{\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_t\mathbf{A}_d & \mathbf{I} \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_t\mathbf{A}_d & \mathbf{I} \end{pmatrix}}_{=\mathbf{I}} \begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}\Bigg] \\
=&(2\pi)^{-\frac{n_p+n_w}{2}} \left|\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_t\mathbf{A}_d & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right|^{-\frac{1}{2}} \\
&\times \exp\Bigg[-\frac{1}{2}\begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{M}_t\mathbf{A}_d\mathbf{d} - \mathbf{M}_t\mathbf{A}_d\boldsymbol{\delta}_t + \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}^T \\
&\qquad \times \left(\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{M}_t\mathbf{A}_d & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_t \end{pmatrix}\begin{pmatrix} \mathbf{I} & \mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{0} & \mathbf{I} \end{pmatrix}\right)^{-1} \\
&\qquad \times \begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{M}_t\mathbf{A}_d\mathbf{d} - \mathbf{M}_t\mathbf{A}_d\boldsymbol{\delta}_t + \mathbf{w}_t - \tilde{\mu}_t - \mathbf{M}_t\mathbf{A}_d\mathbf{d} \end{pmatrix}\Bigg] \\
=&(2\pi)^{-\frac{n_p+n_w}{2}} \left|\begin{pmatrix} \mathbf{D}_t & \mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t & \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T \end{pmatrix}\right|^{-\frac{1}{2}} \\
&\times \exp\Bigg[-\frac{1}{2}\begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \left(\tilde{\mu}_t + \mathbf{M}_t\mathbf{A}_d\boldsymbol{\delta}_t\right) \end{pmatrix}^T \begin{pmatrix} \mathbf{D}_t & \mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T \\ \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t & \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T \end{pmatrix}^{-1} \\
&\qquad \times \begin{pmatrix} \mathbf{d}-\boldsymbol{\delta}_t \\ \mathbf{w}_t - \left(\tilde{\mu}_t + \mathbf{M}_t\mathbf{A}_d\boldsymbol{\delta}_t\right) \end{pmatrix}\Bigg].
\end{aligned}
$$

Consequently, it follows from (D.19) and (D.20) that $\mathbf{w}_t$ given $\mathbf{Y}_t$ is normally distributed with mean vector $\boldsymbol{\mu}_t$ and variance matrix $\mathbf{C}_t$ defined by

$$\boldsymbol{\mu}_t := \mathrm{E}[\mathbf{w}_t|\mathbf{Y}_t] = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\mathbf{A}_d\boldsymbol{\delta}_t,$$
$$\mathbf{C}_t := \mathrm{Var}[\mathbf{w}_t|\mathbf{Y}_t] = \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T.$$

The claim about log-density of $\mathbf{Y}_N$ then follows directly from equations (D.9), (D.11), (D.13) and (D.14):

$$\begin{aligned}
\log(f_{\mathbf{Y}_N}) &= \log(f_{\mathbf{Y}_N|\mathbf{d}}) + \log(f_{\mathbf{d}}) - \log(f_{\mathbf{d}|\mathbf{Y}_N}) \\
&= \underbrace{\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N\ - \frac{1}{2}\ \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\ \mathbf{d}}_{\log(f_{\mathbf{Y}_N|\mathbf{d}}),\text{ from (D.9).}} \\
&\quad \underbrace{-\frac{n_d}{2}\log(2\pi)\ - \frac{1}{2}\log|\mathbf{D}_0|\ - \frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_0)^T\ \mathbf{D}_0^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_0)}_{=\log(f_\mathbf{d}),\text{ from (D.14).}} \\
&\quad -\underbrace{\left(-\frac{n_d}{2}\log(2\pi)\ - \frac{1}{2}\log|\mathbf{D}_N|\ - \frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_N)^T\ \mathbf{D}_N^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_N)\right)}_{=\log(f_{\mathbf{d}|\mathbf{Y}_N}),\text{ from (D.11).}} \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\log|\mathbf{D}_0| + \frac{1}{2}\log|\mathbf{D}_N| + \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N\ - \frac{1}{2}\ \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\ \mathbf{d} \\
&\quad - \frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_0)^T\ \mathbf{D}_0^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_0) \\
&\quad + \frac{1}{2}(\mathbf{d}-\boldsymbol{\delta}_N)^T\ \mathbf{D}_N^{-1}\ (\mathbf{d}-\boldsymbol{\delta}_N) \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}(\log|\mathbf{D}_0| - \log|\mathbf{D}_N|) + \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N\ - \frac{1}{2}\ \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\ \mathbf{d} \\
&\quad - \frac{1}{2}\mathbf{d}^T\mathbf{D}_0^{-1}\mathbf{d}\ \underbrace{+\frac{1}{2}\mathbf{d}^T\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_0^{-1}\mathbf{d}}_{=\mathbf{d}^T\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0\right),\text{ since }\left(\boldsymbol{\delta}_0^T\mathbf{D}_0^{-1}\mathbf{d}\right)\in\mathbb{R}.}\ - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 \\
&\quad + \frac{1}{2}\mathbf{d}^T\mathbf{D}_N^{-1}\mathbf{d}\ \underbrace{-\frac{1}{2}\mathbf{d}^T\mathbf{D}_N^{-1}\boldsymbol{\delta}_N - \frac{1}{2}\boldsymbol{\delta}_N^T\mathbf{D}_N^{-1}\mathbf{d}}_{=-\mathbf{d}^T\left(\mathbf{D}_N^{-1}\boldsymbol{\delta}_N\right),\text{ since }\left(\boldsymbol{\delta}_N^T\mathbf{D}_N^{-1}\mathbf{d}\right)\in\mathbb{R}.}\ + \frac{1}{2}\boldsymbol{\delta}_N^T\mathbf{D}_N^{-1}\boldsymbol{\delta}_N \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\log\left|\mathbf{D}_0\mathbf{D}_N^{-1}\right| - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \frac{1}{2}\boldsymbol{\delta}_N^T\mathbf{D}_N^{-1}\boldsymbol{\delta}_N \\
&\quad - \frac{1}{2}\underbrace{\left(\mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\ \mathbf{d} + \mathbf{d}^T\mathbf{D}_0^{-1}\mathbf{d}\right)}_{=\mathbf{d}^T\left(\mathbf{D}_0^{-1}+\mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\right)\mathbf{d}} + \underbrace{\mathbf{d}^T\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0\right) + \mathbf{d}^T\ \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N}_{=\mathbf{d}^T\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0+\mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N\right)} \\
&\quad + \frac{1}{2}\ \underbrace{\mathbf{d}^T\mathbf{D}_N^{-1}\mathbf{d}}_{=\mathbf{d}^T\left(\mathbf{D}_0^{-1}+\mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\right)\mathbf{d}}\ - \underbrace{\mathbf{d}^T\left(\mathbf{D}_N^{-1}\boldsymbol{\delta}_N\right)}_{=\mathbf{d}^T\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0+\mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N\right)} \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 \\
&\quad - \underbrace{\frac{1}{2}\log\left|\mathbf{D}_0\mathbf{D}_N^{-1}\right|}_{=-\frac{1}{2}\log\left|\mathbf{D}_0\left(\mathbf{D}_0^{-1}+\mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}\right)\right|} + \underbrace{\frac{1}{2}\boldsymbol{\delta}_N^T\mathbf{D}_N^{-1}\boldsymbol{\delta}_N}_{=\frac{1}{2}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0+\mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N)^T(\mathbf{D}_0^{-1}+\mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d})^{-1}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0+\mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N)}
\end{aligned}$$

$$= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) - \frac{1}{2}\log|\mathbf{I} + \mathbf{D}_0\ \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d}| - \frac{1}{2}\boldsymbol{\delta}_0^T\mathbf{D}_\mathbf{0}^{-1}\boldsymbol{\delta}_0$$
$$+ \frac{1}{2}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N)^T(\mathbf{D}_0^{-1} + \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_N\ \mathbf{A}_\mathbf{d})^{-1}(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_N),$$

which completes the proof.[27]

□

Note that (16) follows directly from the claims of Proposition D.2, while we may obtain (15a) and (15b) by substituting $\boldsymbol{\delta}_t = \left(\mathbf{D}_0^{-1} + \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_t\ \mathbf{A}_\mathbf{d}\right)^{-1}\left(\mathbf{D}_0^{-1}\boldsymbol{\delta}_0 + \mathbf{A}_\mathbf{d}^T\ \mathbf{s}_t\right)$ and $\mathbf{D}_t = \left(\mathbf{D}_0^{-1} + \mathbf{A}_\mathbf{d}^T\ \mathbf{S}_t\ \mathbf{A}_\mathbf{d}\right)^{-1}$ into $\boldsymbol{\mu}_t = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\mathbf{A}_d\ \boldsymbol{\delta}_t$ and $\mathbf{C}_t = \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d\mathbf{D}_t\mathbf{A}_d^T\mathbf{M}_t^T$, respectively. To derive the remaining formulas of the AKF, notice that defining the sequences $\{\mathbf{s}_t\}_{t=0}^N$, $\{\mathbf{M}_t\}_{t=0}^N$, and $\{\mathbf{S}_t\}_{t=0}^N$ as in Propositions D.1 and D.2 is equivalent to their recursive derivation given in (15c)-(15e). To see this, note that

$$\begin{aligned}
\mathbf{s}_t &= \sum_{i=1}^{t}\mathbf{E}_i^T\tilde{\mathbf{U}}_i^{-1}\ \tilde{\mathbf{e}}_i \\
&= \sum_{i=1}^{t}(\mathbf{HFM}_{i-1})^T\tilde{\mathbf{U}}_i^{-1}\ \tilde{\mathbf{e}}_i \\
&= (\mathbf{HFM}_{t-1})^T\tilde{\mathbf{U}}_t^{-1}\ \tilde{\mathbf{e}}_t + \sum_{i=1}^{t-1}(\mathbf{HFM}_{i-1})^T\tilde{\mathbf{U}}_i^{-1}\ \tilde{\mathbf{e}}_i \\
&= \mathbf{s}_{t-1} + (\mathbf{HFM}_{t-1})^T\tilde{\mathbf{U}}_t^{-1}\ \tilde{\mathbf{e}}_t, \\
\mathbf{S}_t &= \sum_{i=1}^{t}\mathbf{E}_i^T\tilde{\mathbf{U}}_i^{-1}\ \mathbf{E}_i \\
&= \sum_{i=1}^{t}(\mathbf{HFM}_{i-1})^T\tilde{\mathbf{U}}_i^{-1}\ (\mathbf{HFM}_{i-1}) \\
&= (\mathbf{HFM}_{t-1})^T\tilde{\mathbf{U}}_t^{-1}\ (\mathbf{HFM}_{t-1}) + \sum_{i=1}^{t-1}(\mathbf{HFM}_{i-1})^T\tilde{\mathbf{U}}_i^{-1}\ (\mathbf{HFM}_{i-1}) \\
&= \mathbf{S}_{t-1} + (\mathbf{HFM}_{t-1})^T\tilde{\mathbf{U}}_t^{-1}\ (\mathbf{HFM}_{t-1}), \\
\mathbf{M}_t &= \prod\nolimits_{\mathbf{j=1}}^{\mathbf{t}}\mathbf{J}_{t+1-j} \\
&= \mathbf{J}_t\ \mathbf{J}_{t-1}\ \ldots\ \mathbf{J}_1 \\
&= \mathbf{J}_t\ \prod\nolimits_{\mathbf{j=1}}^{\mathbf{t-1}}\mathbf{J}_{t-j} \\
&= \mathbf{J}_t\ \mathbf{M}_{t-1} \\
&= \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{F}\ \mathbf{M}_{t-1}.
\end{aligned}$$

[27]Note that some arguments of this proof are taken from Durbin and Koopman (2012, pp. 141-144) and de Jong (1988).

Further, by definition of the $\sum(\cdot)$ and the $\prod(\cdot)$ operator, we get:

$$\mathbf{s}_0 = \sum_{i=1}^{0} \mathbf{E}_i^T \tilde{\mathbf{U}}_i^{-1} \, \tilde{\mathbf{e}}_i = \mathbf{0}, \qquad \mathbf{S}_0 = \sum_{i=1}^{0} \mathbf{E}_i^T \tilde{\mathbf{U}}_i^{-1} \, \mathbf{E}_i = \mathbf{0}, \qquad \mathbf{M}_0 = \prod_{\mathbf{j=1}}^{0} \mathbf{J}_{1-j} = \mathbf{I}.$$

D.2 Initialization strategies for non-stationary state-space models

In the following, we present two well-known strategies, namely the *fixed-but-unknown* and the *diffuse* initialization, to choose $(\boldsymbol{\mu}_0, \mathbf{C}_0)$ in the context of non-stationary SSMs.

Fixed-but-unknown initialization: Imagine the state vector $\mathbf{w}_t$ contains some non-stationary elements, which implies that the unconditional second moments of $\mathbf{w}_t$ do not exist; therefore an unconditional initialization is impossible. One way to handle non-stationary SSMs is to treat the non-stationary elements in $\mathbf{w}_0$ as fixed-but-unknown and estimate them via maximum-likelihood. We refer to this approach, which goes back to Rosenberg (1973), as the *fixed-but-unknown* initialization. As de Jong (1988) shows, we can easily apply the *fixed-but-unknown* initialization within the framework of the AKF. To see this, suppose that we may reorder the initial state vector $\mathbf{w}_0$ such that

$$\mathbf{w}_0 = \begin{pmatrix} \mathbf{w}_0^{(1)} \\ \mathbf{w}_0^{(2)} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_0^{(1)} \\ \boldsymbol{\mu}_0^{(2)} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_0^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_0^{(2)} \end{pmatrix} \right), \quad \boldsymbol{\mu}_0^{(1)} = \boldsymbol{\mu}^{(1)}, \quad \mathbf{C}_0^{(1)} = \mathbf{C}^{(1)}, \quad \mathbf{C}_0^{(2)} = z\mathbf{I}, \qquad \text{(D.21)}$$

where $\mathbf{w}_0^{(1)}$ and $\mathbf{w}_0^{(2)}$ denote the stationary and non-stationary elements, respectively, and where $\boldsymbol{\mu}^{(1)}$ and $\mathbf{C}^{(1)}$ represent the unconditional mean vector and the unconditional variance matrix of $\mathbf{w}_t^{(1)}$. For an initial state vector $\mathbf{w}_0$ as defined in (D.21), we can consider the fixed-but-unknown initialization as the case where z tends to zero. We may also express $\mathbf{w}_0$, defined via (D.21) using (13) by setting

$$\bar{\boldsymbol{\mu}}_0 = \mathbf{0}, \qquad \bar{\mathbf{C}}_0 = \mathbf{C}_0^{(1)}, \qquad \mathbf{a}_\mathbf{w} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \qquad \mathbf{A}_\mathbf{w} = \begin{pmatrix} \mathbf{I} \\ \mathbf{0} \end{pmatrix},$$

$$\boldsymbol{\delta}_0 = \boldsymbol{\mu}_0^{(2)} \qquad \mathbf{D}_0 = \mathbf{C}_0^{(2)} = z\mathbf{I}, \qquad \mathbf{A}_\mathbf{d} = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}.$$

Hence, we can apply the AKF to the initial state vector $\mathbf{w}_0$ defined by (D.21). If we now let z tend to zero, i.e., $\mathbf{d} = \boldsymbol{\delta}_0$ and $\mathbf{D}_0 \to \mathbf{0}$, (15a) and (15b) become

$$\begin{aligned} \boldsymbol{\mu}_{t|\mathbf{d}} &= \mathrm{E}[\mathbf{w}_t | \mathbf{Y}_t, \mathbf{d}] \\ &= \lim_{\mathbf{D}_0 \to \mathbf{0}} \boldsymbol{\mu}_t \\ &= \lim_{\mathbf{D}_0 \to \mathbf{0}} \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t \mathbf{A}_d \left(\mathbf{D}_0^{-1} + \mathbf{A}_\mathbf{d}^T \, \mathbf{S}_t \, \mathbf{A}_\mathbf{d} \right)^{-1} \left(\mathbf{D}_0^{-1} \boldsymbol{\delta}_0 + \mathbf{A}_\mathbf{d}^T \, \mathbf{s}_t \right) \end{aligned}$$

$$
\begin{aligned}
&= \lim_{\mathbf{D}_0 \to \mathbf{0}} \tilde{\mu}_t + \mathbf{M}_t \mathbf{A}_d \left(\mathbf{I} + \mathbf{D}_0 \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_t \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \mathbf{D}_0 \left(\mathbf{D}_0^{-1} \boldsymbol{\delta}_0 + \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_t\right) \\
&= \lim_{\mathbf{D}_0 \to \mathbf{0}} \tilde{\mu}_t + \mathbf{M}_t \mathbf{A}_d \left(\mathbf{I} + \mathbf{D}_0 \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_t \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \left(\boldsymbol{\delta}_0 + \mathbf{D}_0 \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_t\right) \\
&= \tilde{\mu}_t + \mathbf{M}_t \mathbf{A}_d \ (\mathbf{I})^{-1} \mathbf{d} \\
&= \tilde{\mu}_t + \mathbf{M}_t \mathbf{A}_d \ \mathbf{d}, \qquad \forall\ t = 1, 2, \ldots, N, \qquad \text{(D.22a)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{C}_{t|\mathbf{d}} &= \mathrm{Var}[\mathbf{w}_t | \mathbf{Y}_t, \mathbf{d}] \\
&= \lim_{\mathbf{D}_0 \to \mathbf{0}} \mathbf{C}_t \\
&= \lim_{\mathbf{D}_0 \to \mathbf{0}} \tilde{\mathbf{C}}_t + \mathbf{M}_t \mathbf{A}_d \left(\mathbf{D}_0^{-1} + \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_t \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \mathbf{A}_d^T \mathbf{M}_t^T \\
&= \lim_{\mathbf{D}_0 \to \mathbf{0}} \tilde{\mathbf{C}}_t + \mathbf{M}_t \mathbf{A}_d \left(\mathbf{I} + \mathbf{D}_0 \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_t \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \mathbf{D}_0 \mathbf{A}_d^T \mathbf{M}_t^T \\
&= \tilde{\mathbf{C}}_t, \qquad \forall\ t = 1, 2, \ldots, N, \qquad \text{(D.22b)}
\end{aligned}
$$

and the conditional log-density of $\mathbf{Y}_N$ given $\mathbf{d}$ follows from Proposition D.1 as

$$
\log\left(f_{\mathbf{Y}_N|\mathbf{d}}\right) = \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \mathbf{d}^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \ - \frac{1}{2} \ \mathbf{d}^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}} \ \mathbf{d}. \qquad \text{(D.23)}
$$

Further, if $\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}$ is non-singular, we may obtain the maximum-likelihood estimator of $\mathbf{d}$ and its estimated variance matrix from Proposition D.1 as

$$
\begin{aligned}
\hat{\mathbf{d}} &= \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N, \qquad \text{(D.24a)} \\
\widehat{\mathrm{Var}}[\hat{\mathbf{d}}] &= \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1}. \qquad \text{(D.24b)}
\end{aligned}
$$

Hence, substituting $\mathbf{d}$ by $\hat{\mathbf{d}}$ in (D.23) yields

$$
\begin{aligned}
\log\left(f_{\mathbf{Y}_N|\mathbf{d}=\hat{\mathbf{d}}}\right) &= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \hat{\mathbf{d}}^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \ - \frac{1}{2} \ \hat{\mathbf{d}}^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}} \ \hat{\mathbf{d}} \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \left[\left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N\right]^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \\
&\quad - \frac{1}{2} \left[\left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N\right]^T \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}} \left[\left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N\right] \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \mathbf{s}_N^T \ \mathbf{A}_{\mathbf{d}} \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \\
&\quad - \frac{1}{2} \ \mathbf{s}_N^T \ \mathbf{A}_{\mathbf{d}} \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}} \ \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \mathbf{s}_N^T \ \mathbf{A}_{\mathbf{d}} \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N - \frac{1}{2} \ \mathbf{s}_N^T \ \mathbf{A}_{\mathbf{d}} \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N \\
&= \log\left(f_{\mathbf{Y}_N|\mathbf{d}=\mathbf{0}}\right) + \frac{1}{2} \ \mathbf{s}_N^T \ \mathbf{A}_{\mathbf{d}} \ \left(\mathbf{A}_{\mathbf{d}}^T \ \mathbf{S}_N \ \mathbf{A}_{\mathbf{d}}\right)^{-1} \ \mathbf{A}_{\mathbf{d}}^T \ \mathbf{s}_N, \qquad \text{(D.25)}
\end{aligned}
$$

which is the log-density of $\mathbf{Y}_N$ concentrated with respect to $\mathbf{d}$. This means that if we want to evaluate the log-density of $\mathbf{Y}_N$ based on the *fixed-but-unknown* initialization for the non-stationary elements $\mathbf{w}_0^{(2)}$ of $\mathbf{w}_0$, we replace (6) with (D.25).

Diffuse initialization: A rather contrary approach to the *fixed-but-unknown* initialization is the so-called *diffuse* initialization, where we treat the non-stationary elements $\mathbf{w}_0^{(2)}$ of the initial state vector $\mathbf{w}_0$ as diffuse, i.e., $z \to \infty$ or equivalently $\mathbf{D}_0 \to \infty$. Thus, treating $\mathbf{w}_0^{(2)}$ as diffuse, (15a) and (15b) yield

$$\boldsymbol{\mu}_{t,z\to\infty} := \lim_{z\to\infty} \boldsymbol{\mu}_t = \tilde{\boldsymbol{\mu}}_t + \mathbf{M}_t\mathbf{A}_d \left(\mathbf{A}_\mathbf{d}^T \ \mathbf{S}_t \ \mathbf{A}_\mathbf{d}\right)^{-1} \left(\mathbf{A}_\mathbf{d}^T \ \mathbf{s}_t\right), \qquad \forall\ t = k, k+1, \dots, N, \tag{D.26a}$$

$$\mathbf{C}_{t,z\to\infty} := \lim_{z\to\infty} \mathbf{C}_t = \tilde{\mathbf{C}}_t + \mathbf{M}_t\mathbf{A}_d \left(\mathbf{A}_\mathbf{d}^T \ \mathbf{S}_t \ \mathbf{A}_\mathbf{d}\right)^{-1} \mathbf{A}_d^T\mathbf{M}_t^T, \qquad \forall\ t = k, k+1, \dots, N, \tag{D.26b}$$

where k is the first period where the matrix $\mathbf{A}_\mathbf{d}^T \ \mathbf{S}_t \ \mathbf{A}_\mathbf{d}$ becomes non-singular. Thus, although the initial state vector $\mathbf{w}_0$ in this case ($z \to \infty$) has an improper distribution, in the sense that it does not integrate to one, it has a proper distribution conditional on $\mathbf{Y}_k$. In practice, when dealing with non-stationary SSMs, we often use often only the first k observations to obtain $\boldsymbol{\mu}_{k,z\to\infty}$ and $\mathbf{C}_{k,z\to\infty}$. We then use the remaining observations to evaluate the log-likelihood based on the original Kalman recursion (4) initialized at $(\boldsymbol{\mu}_{k,z\to\infty}, \mathbf{C}_{k,z\to\infty})$. For more detailed treatments of the diffuse initialization using the AKF, we refer the reader to textbook treatments by Harvey (1990b) or Durbin and Koopman (2012).

D.3 Incorporating the augmented Kalman filter into the Kalman recursion

To compute the log-likelihood $\log\left(f_{\mathbf{Y}_N}\right)$ based on the AKF, one augments the standard Kalman recursion (4) initialized at $(\tilde{\boldsymbol{\mu}}_0, \tilde{\mathbf{C}}_0)$ so that for all $t = 1, 2, \dots, N$, the quantity $\mathbf{M}_t$ can be computed in parallel. To do so, we define

$$\mathbf{W}_{t|t-1} := \mathbf{F}\mathbf{M}_{t-1}, \qquad ,\forall\ t = 1, 2, \dots, N \tag{D.27}$$

so that we may compute $\mathbf{M}_t$ as

$$\begin{aligned}
\mathbf{M}_t &= \prod\nolimits_{\mathrm{j=1}}^{\mathrm{t}} \mathbf{J}_{t-j+1} \\
&= \mathbf{J}_t\mathbf{M}_{t-1} \\
&= \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{F}\mathbf{M}_{t-1} \\
&= \left(\mathbf{I} - \tilde{\mathbf{K}}_t\mathbf{H}\right)\mathbf{W}_{t|t-1} \\
&= \mathbf{W}_{t|t-1} - \tilde{\mathbf{K}}_t\mathbf{H}\mathbf{W}_{t|t-1} \\
&= \mathbf{W}_{t|t-1} - \tilde{\mathbf{K}}_t\mathbf{E}_t \qquad ,\forall\ t = 1, 2, \dots, N.
\end{aligned} \tag{D.28}$$

Based on (D.27) and (D.28), we get the augmented Kalman recursion by extending (4a), (4c), and (4e) from the standard Kalman recursion (4) initialized at $(\tilde{\mu}_0, \tilde{\mathbf{C}}_0)$ to

$$\begin{pmatrix} \tilde{\mathbf{w}}_{t|t-1} & \mathbf{W}_{t|t-1} \end{pmatrix} = \mathbf{F} \begin{pmatrix} \tilde{\mu}_{t-1} & \mathbf{M}_{t-1} \end{pmatrix}, \tag{D.29a}$$

$$\begin{pmatrix} \tilde{\mathbf{e}}_t & \mathbf{E}_t \end{pmatrix} = \begin{pmatrix} \mathbf{y}_t^{(\mathbf{h})} & \mathbf{0} \end{pmatrix} - \mathbf{H} \begin{pmatrix} \tilde{\mathbf{w}}_{t|t-1} & \mathbf{W}_{t|t-1} \end{pmatrix}, \tag{D.29c}$$

$$\begin{pmatrix} \tilde{\mu}_t & \mathbf{M}_t \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{w}}_{t|t-1} & \mathbf{W}_{t|t-1} \end{pmatrix} + \tilde{\mathbf{K}}_t \begin{pmatrix} \tilde{\mathbf{e}}_t & \mathbf{E}_t \end{pmatrix}, \tag{D.29e}$$

for all $t = 1, 2, \ldots, N$.

E Smets and Wouters model

The version of the dynamic stochastic general equilibrium (DSGE) model introduced by Smets and Wouters (2007) that we use in this paper reflects a slightly adjusted version of the original model that follows the Dynare implementation made available by Johannes Pfeifer.[28]

E.1 Stochastic process and residuals

The stochastic process driving the economy is given by

$$\begin{aligned}
\varepsilon_t^a &= \rho_a \, \varepsilon_{t-1}^a + \eta_t^a & \eta_t^a &\sim N(0, \sigma_a),\\
\varepsilon_t^b &= \rho_b \, \varepsilon_{t-1}^b + \eta_t^b & \eta_t^b &\sim N(0, \sigma_b),\\
\varepsilon_t^g &= \rho_g \, \varepsilon_{t-1}^g + \rho_{ga} \, \eta_t^a + \eta_t^g & \eta_t^g &\sim N(0, \sigma_g),\\
\varepsilon_t^i &= \rho_i \, \varepsilon_{t-1}^i + \eta_t^i & \eta_t^i &\sim N(0, \sigma_i),\\
\varepsilon_t^r &= \rho_r \, \varepsilon_{t-1}^r + \eta_t^r & \eta_t^r &\sim N(0, \sigma_r),\\
\varepsilon_t^p &= \rho_p \, \varepsilon_{t-1}^p - \mu_p \, \eta_{t-1}^p + \eta_t^p & \eta_t^p &\sim N(0, \sigma_p),\\
\varepsilon_t^w &= \rho_w \, \varepsilon_{t-1}^w - \mu_w \, \eta_{t-1}^w + \eta_t^w & \eta_t^w &\sim N(0, \sigma_w),
\end{aligned}$$

where ε_t^a, ε_t^b, ε_t^g, ε_t^i, ε_t^r, ε_t^p, and ε_t^w denote a productivity shock, a risk premium shock, an exogenous government spending shock, an investment-specific technology shock, a monetary policy shock, a price markup shock, and a wage markup shock, respectively. In two periods, we can write this stochastic process as

$$0 = \varepsilon_t^a - \eta_t^a - \rho_a \, L\left(\varepsilon_t^a\right), \tag{E.1}$$

$$0 = \varepsilon_t^b - \eta_t^b - \rho_b \, L\left(\varepsilon_t^b\right), \tag{E.2}$$

$$0 = \varepsilon_t^g - \eta_t^g - \eta_t^a \, \rho_{ga} - \rho_g \, L\left(\varepsilon_t^g\right), \tag{E.3}$$

[28] Link: https://github.com/JohannesPfeifer/DSGE_mod/blob/master/Smets_Wouters_2007/Smets_Wouters_2007_45.mod

$$0 = \varepsilon_t^i - \eta_t^i - \rho_i\, L\left(\varepsilon_t^i\right), \tag{E.4}$$

$$0 = \varepsilon_t^r - \eta_t^r - \rho_r\, L\left(\varepsilon_t^r\right), \tag{E.5}$$

$$0 = \varepsilon_t^p - \eta_t^p + L\left(\eta_t^p\right)\mu_p - \rho_p\, L\left(\varepsilon_t^p\right), \tag{E.6}$$

$$0 = \varepsilon_t^w - \eta_t^w + L\left(\eta_t^w\right)\mu_w - \rho_w\, L\left(\varepsilon_t^w\right), \tag{E.7}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^a\right], \tag{E.8}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^b\right], \tag{E.9}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^g\right], \tag{E.10}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^i\right], \tag{E.11}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^r\right], \tag{E.12}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^p\right], \tag{E.13}$$

$$0 = \mathbb{E}_t\left[\eta_{t+1}^w\right], \tag{E.14}$$

where $L(x_t)$ denotes the variable x_t lagged by one period, i.e, $L(x_t) = x_{t-1}$.

E.2 Economy with sticky prices and wages

At the core of the log-linearized version of the model are 13 equations

$$0 = y_t - \Phi\,\varepsilon_t^a - \alpha\,\Phi\,k_t^s + \Phi\,l_t\,(\alpha - 1), \tag{E.15}$$

$$0 = k_t^s - L\left(k_t\right) - z_t, \tag{E.16}$$

$$0 = z_t + \frac{r_t^k\,(\psi - 1)}{\psi}, \tag{E.17}$$

$$0 = \mu_{p\,t} + \varepsilon_t^a - \alpha\, r_t^k + w_t\,(\alpha - 1), \tag{E.18}$$

$$0 = k_t^s - l_t + r_t^k - w_t, \tag{E.19}$$

$$0 = y_t - \varepsilon_t^g - c_t\,c_y - i_t\,i_y - z_t\,z_y, \tag{E.20}$$

$$\begin{aligned} 0 = {} & r_t - \varepsilon_t^r - L\left(r_t\right)\rho + r_{\Delta y}\,L\left(y_t\right) - r_{\Delta y}\,L\left(y_t^f\right) - y_t\left(r_{\Delta y} - r_y\,(\rho - 1)\right) \\ & + y_t^f\left(r_{\Delta y} - r_y\,(\rho - 1)\right) + \pi_t\,r_\pi\,(\rho - 1), \end{aligned} \tag{E.21}$$

$$0 = -i_{k,\gamma}\,\varepsilon_t^i\,\varphi\,\gamma^2 + k_t - i_t\,i_{k,\gamma} + L\left(k_t\right)\left(i_{k,\gamma} - 1\right), \tag{E.22}$$

$$0 = i_t - \varepsilon_t^i - \frac{L\left(i_t\right)}{\bar\beta\,\gamma + 1} - \frac{q_t}{\gamma^2\,\varphi\,\left(\bar\beta\,\gamma + 1\right)} - \frac{\mathbb{E}_t\left[i_{t+1}\right]\bar\beta\,\gamma}{\bar\beta\,\gamma + 1}, \tag{E.23}$$

$$0 = q_t - \mathbb{E}_t\left[\pi_{t+1}\right] + r_t - \frac{\mathbb{E}_t\left[r_{t+1}^k\right] r_{ss}^k}{r_{ss}^k - \delta + 1} + \frac{\mathbb{E}_t\left[q_{t+1}\right](\delta - 1)}{r_{ss}^k - \delta + 1} + \frac{\sigma_c\,\varepsilon_t^b\left(\frac{h}{\gamma} + 1\right)}{\frac{h}{\gamma} - 1}, \tag{E.24}$$

$$0 = c_t - \varepsilon_t^b - \frac{\mathbb{E}_t\left[c_{t+1}\right]}{\frac{h}{\gamma} + 1} - \frac{L\left(c_t\right)h}{\gamma\left(\frac{h}{\gamma} + 1\right)} + \frac{\mathbb{E}_t\left[\pi_{t+1}\right]\left(\frac{h}{\gamma} - 1\right)}{\sigma_c\left(\frac{h}{\gamma} + 1\right)} - \frac{r_t\left(\frac{h}{\gamma} - 1\right)}{\sigma_c\left(\frac{h}{\gamma} + 1\right)}$$

$$+\frac{\mathbb{E}_t\left[l_{t+1}\right] wl_c\,(\sigma_c-1)}{\sigma_c\left(\frac{h}{\gamma}+1\right)}-\frac{l_t\, wl_c\,(\sigma_c-1)}{\sigma_c\left(\frac{h}{\gamma}+1\right)}, \tag{E.25}$$

$$0=\pi_t-\varepsilon_t^p-\frac{\iota_p\, L(\pi_t)}{\bar\beta\,\gamma\,\iota_p+1}-\frac{\mathbb{E}_t\left[\pi_{t+1}\right]\bar\beta\,\gamma}{\bar\beta\,\gamma\,\iota_p+1}-\frac{\mu_{p_t}\left(\bar\beta\,\gamma\,\xi_p-1\right)\left(\xi_p-1\right)}{\xi_p\left(\varepsilon_p\,(\Phi-1)+1\right)\left(\bar\beta\,\gamma\,\iota_p+1\right)}, \tag{E.26}$$

$$\begin{aligned}0=w_t&\left(\frac{\left(\bar\beta\,\gamma\,\xi_w-1\right)(\xi_w-1)}{\xi_w\left(\bar\beta\,\gamma+1\right)\left(\varepsilon_w\,(\lambda_w-1)+1\right)}+1\right)-\varepsilon_t^w-\frac{L(w_t)}{\bar\beta\,\gamma+1}-\frac{\iota_w\, L(\pi_t)}{\bar\beta\,\gamma+1}\\&+\frac{\pi_t\left(\bar\beta\,\gamma\,\iota_w+1\right)}{\bar\beta\,\gamma+1}-\frac{\mathbb{E}_t\left[\pi_{t+1}\right]\bar\beta\,\gamma}{\bar\beta\,\gamma+1}-\frac{\mathbb{E}_t\left[w_{t+1}\right]\bar\beta\,\gamma}{\bar\beta\,\gamma+1}\\&-\frac{l_t\,\sigma_l\left(\bar\beta\,\gamma\,\xi_w-1\right)(\xi_w-1)}{\xi_w\left(\bar\beta\,\gamma+1\right)\left(\varepsilon_w\,(\lambda_w-1)+1\right)}+\frac{c_t\left(\bar\beta\,\gamma\,\xi_w-1\right)(\xi_w-1)}{\xi_w\left(\bar\beta\,\gamma+1\right)\left(\frac{h}{\gamma}-1\right)\left(\varepsilon_w\,(\lambda_w-1)+1\right)}\\&-\frac{L(c_t)\,h\left(\bar\beta\,\gamma\,\xi_w-1\right)(\xi_w-1)}{\gamma\,\xi_w\left(\bar\beta\,\gamma+1\right)\left(\frac{h}{\gamma}-1\right)\left(\varepsilon_w\,(\lambda_w-1)+1\right)},\end{aligned} \tag{E.27}$$

in the 14 endogenous variables that describe an economy with sticky price and wage contracts: output y_t, consumption c_t, investment i_t, hours worked l_t, capital services k_t^s, capital stock k_t, real wage w_t, rental rate of capital r_t^k, capital utilization rate z_t, real value of existing capital stock q_t, inflation π_t, nominal interest rate r_t, gross price markup μ_{p_t}, and potential output y_t^f.

E.3 Economy with flexible prices and wages

To determine potential output y_t^f the model is augmented by the 11 equations

$$0=y_t^f-\Phi\,\varepsilon_t^a-\alpha\,\Phi\, k_t^{s,f}+\Phi\, l_t^f\,(\alpha-1), \tag{E.28}$$

$$0=k_t^{s,f}-L\left(k_t^f\right)-z_t^f, \tag{E.29}$$

$$0=z_t^f+\frac{r_t^{kf}\,(\psi-1)}{\psi}, \tag{E.30}$$

$$0=\varepsilon_t^a-\alpha\, r_t^{kf}+w_t^f\,(\alpha-1), \tag{E.31}$$

$$0=k_t^{s,f}-l_t^f+r_t^{kf}-w_t^f, \tag{E.32}$$

$$0=w_t^f+\frac{c_t^f}{\frac{h}{\gamma}-1}-l_t^f\,\sigma_l-\frac{L\left(c_t^f\right)h}{\gamma\left(\frac{h}{\gamma}-1\right)}, \tag{E.33}$$

$$0=y_t^f-\varepsilon_t^g-c_y\, c_t^f-i_y\, i_t^f-z_y\, z_t^f, \tag{E.34}$$

$$0=-i_{k,\gamma}\,\varepsilon_t^i\,\varphi\,\gamma^2+k_t^f-i_{k,\gamma}\, i_t^f+L\left(k_t^f\right)\left(i_{k,\gamma}-1\right), \tag{E.35}$$

$$0=i_t^f-\varepsilon_t^i-\frac{L\left(i_t^f\right)}{\bar\beta\,\gamma+1}-\frac{q_t^f}{\gamma^2\,\varphi\left(\bar\beta\,\gamma+1\right)}-\frac{\mathbb{E}_t\left[i_{t+1}^f\right]\bar\beta\,\gamma}{\bar\beta\,\gamma+1}, \tag{E.36}$$

$$0=q_t^f+r_t^f-\frac{\mathbb{E}_t\left[r_{t+1}^{kf}\right] r_{ss}^k}{r_{ss}^k-\delta+1}+\frac{\mathbb{E}_t\left[q_{t+1}^f\right](\delta-1)}{r_{ss}^k-\delta+1}+\frac{\sigma_c\,\varepsilon_t^b\left(\frac{h}{\gamma}+1\right)}{\frac{h}{\gamma}-1}, \tag{E.37}$$

$$0 = c_t^f - \varepsilon_t^b - \frac{\mathbb{E}_t\left[c_{t+1}^f\right]}{\frac{h}{\gamma}+1} - \frac{L\left(c_t^f\right)h}{\gamma\left(\frac{h}{\gamma}+1\right)} - \frac{r_t^f\left(\frac{h}{\gamma}-1\right)}{\sigma_c\left(\frac{h}{\gamma}+1\right)} + \frac{\mathbb{E}_t\left[l_{t+1}^f\right]wl_c\left(\sigma_c-1\right)}{\sigma_c\left(\frac{h}{\gamma}+1\right)} - \frac{l_t^f\,wl_c\left(\sigma_c-1\right)}{\sigma_c\left(\frac{h}{\gamma}+1\right)}, \tag{E.38}$$

in the variables y_t^f, c_t^f, i_t^f, l_t^f, $k_t^{s,f}$, k_t^f, w_t^f, $r_t^{k,f}$, z_t^f, q_t^f, and r_t^f, describing the corresponding economy with flexible prices and wages.

E.4 Law of motion of lagged variables

The motion of the model's 20 lagged and therefore predetermined variables is given by

$$0 = L\left(y_{t+1}\right) - y_t, \tag{E.39}$$
$$0 = L\left(c_{t+1}\right) - c_t, \tag{E.40}$$
$$0 = L\left(i_{t+1}\right) - i_t, \tag{E.41}$$
$$0 = L\left(k_{t+1}\right) - k_t, \tag{E.42}$$
$$0 = L\left(r_{t+1}\right) - r_t, \tag{E.43}$$
$$0 = L\left(w_{t+1}\right) - w_t, \tag{E.44}$$
$$0 = L\left(\pi_{t+1}\right) - \pi_t, \tag{E.45}$$
$$0 = L\left(\varepsilon_{t+1}^a\right) - \varepsilon_t^a, \tag{E.46}$$
$$0 = L\left(\varepsilon_{t+1}^b\right) - \varepsilon_t^b, \tag{E.47}$$
$$0 = L\left(\varepsilon_{t+1}^g\right) - \varepsilon_t^g, \tag{E.48}$$
$$0 = L\left(\varepsilon_{t+1}^i\right) - \varepsilon_t^i, \tag{E.49}$$
$$0 = L\left(\varepsilon_{t+1}^r\right) - \varepsilon_t^r, \tag{E.50}$$
$$0 = L\left(\varepsilon_{t+1}^p\right) - \varepsilon_t^p, \tag{E.51}$$
$$0 = L\left(\varepsilon_{t+1}^w\right) - \varepsilon_t^w, \tag{E.52}$$
$$0 = L\left(\eta_{t+1}^p\right) - \eta_t^p, \tag{E.53}$$
$$0 = L\left(\eta_{t+1}^w\right) - \eta_t^w, \tag{E.54}$$
$$0 = L\left(y_{t+1}^f\right) - y_t^f, \tag{E.55}$$
$$0 = L\left(c_{t+1}^f\right) - c_t^f, \tag{E.56}$$
$$0 = L\left(i_{t+1}^f\right) - i_t^f, \tag{E.57}$$
$$0 = L\left(k_{t+1}^f\right) - k_t^f. \tag{E.58}$$

E.5 Data and auxiliary variables

We fit the model to 7 quarterly time series of the log difference of per capita real GDP ($dlGDP_t$), the log difference of per capita real consumption ($dlCONS_t$), the log difference of per capita real investment ($dlINV_t$) and the log difference of per capita real wages ($dlWAGE_t$), log of per capita hours worked ($lHOURs_t/100$), the log difference of GDP deflator (dlP_t), and the federal funds rate ($FEDFUNDS_t$) for U.S. from 1966 to 2004. The series are displayed in Figure 2. To link the model's variables to the data we add 4 auxiliary variables $\bar{y}_t$, $\bar{c}_t$, $\bar{i}_t$, and $\bar{w}_t$ which are determined by

$$0 = \bar{y}_t + L(y_t) - y_t, \tag{E.59}$$

$$0 = \bar{c}_t + L(c_t) - c_t, \tag{E.60}$$

$$0 = \bar{i}_t + L(i_t) - i_t, \tag{E.61}$$

$$0 = \bar{w}_t + L(w_t) - w_t. \tag{E.62}$$

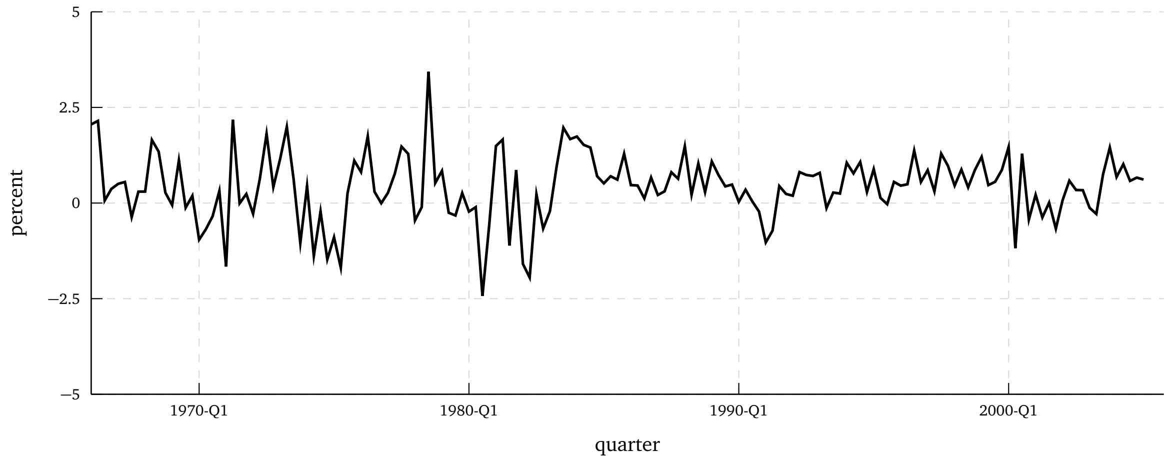
The link between the vector of observations $\mathbf{y}_t$ and the model's variables is then given by

$$\mathbf{y}_t = \begin{pmatrix} dlGDP_t \\ dlCONS_t \\ dlINV_t \\ dlWAGE_t \\ lHOURS_t \\ dlP_t \\ FEDFUNDS_t \end{pmatrix} = \begin{pmatrix} \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{\gamma} \\ \bar{l} \\ \bar{\pi} \\ \bar{r} \end{pmatrix} + \begin{pmatrix} \bar{y}_t \\ \bar{c}_t \\ \bar{i}_t \\ \bar{w}_t \\ l_t \\ \pi_t \\ r_t \end{pmatrix}.$$

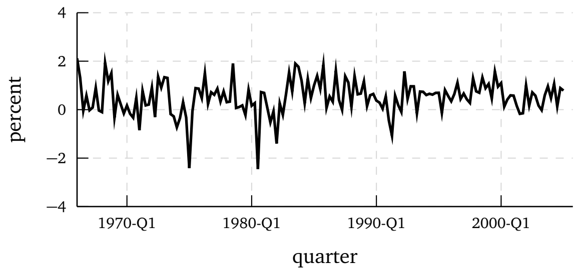
E.6 Policy function and BA-model

The economy described in equations (E.1)-(E.62) includes 62 variables: The 35 endogenous variables y_t, c_t, i_t, l_t, k_t^s, k_t, w_t, r_t^k, z_t, q_t, π_t, r_t, μ_{p_t}, y_t^f, c_t^f, i_t^f, l_t^f, $k_t^{s,f}$, k_t^f, w_t^f, $r_t^{k,f}$, z_t^f, q_t^f, r_t^f, $\bar{y}_t$, $\bar{c}_t$, $\bar{i}_t$, $\bar{w}_t$, ε_t^a, ε_t^b, ε_t^g, ε_t^i, ε_t^r, ε_t^p, and ε_t^w, the 20 predetermined states $L(y_t)$, $L(c_t)$, $L(i_t)$, $L(k_t)$, $L(r_t)$, $L(w_t)$, $L(\pi_t)$, $L(\varepsilon_t^a)$, $L(\varepsilon_t^b)$, $L(\varepsilon_t^g)$, $L(\varepsilon_t^i)$, $L(\varepsilon_t^r)$, $L(\varepsilon_t^p)$, $L(\varepsilon_t^w)$, $L(\eta_t^p)$, $L(\eta_t^w)$, $L(y_t^f)$, $L(c_t^f)$, $L(i_t^f)$, and $L(k_t^f)$, and the 7 exogenous state variables η_t^a, η_t^b, η_t^g, η_t^i, η_t^r, η_t^p, and η_t^w. To solve the model for its policy function we collect the endogenous variables in the vector $\mathbf{y}_t^{(m)}$, the predetermined states in the vector $\mathbf{x}_t^{(m)}$, and the exogenous states in the vector $\mathbf{z}_t^{(m)}$. Since equations (E.1)-(E.62) are linear in $\mathbf{x}_t^{(m)}$, $\mathbf{y}_t^{(m)}$, $\mathbf{z}_t^{(m)}$, $\mathbf{x}_{t+1}^{(m)}$, $\mathbb{E}_t\mathbf{y}_{t+1}^{(m)}$, and $\mathbb{E}_t\mathbf{z}_{t+1}^{(m)}$ they

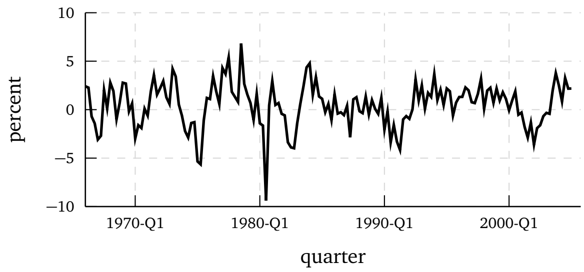
Figure 2: Data – Smets and Wouters model



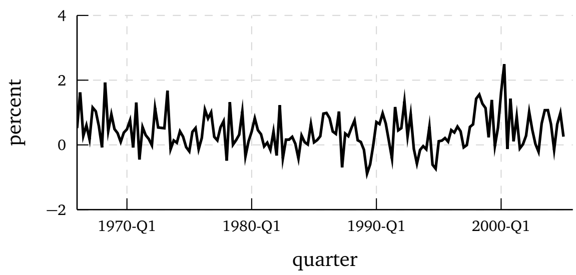
(a) Per Capita Real Output Growth – $dlGDP_t$



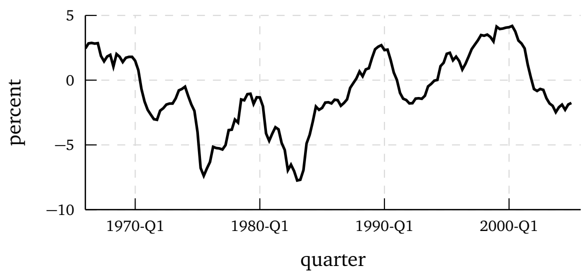
(b) Per Capita Real Consumption Growth – $dlCONS_t$



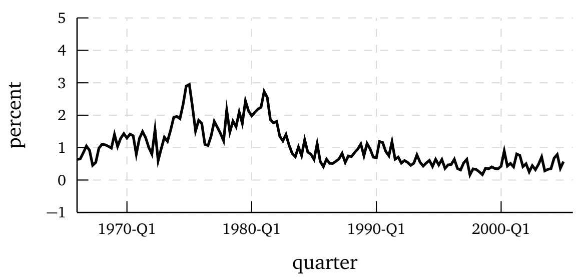
(c) Per Capita Real Investment Growth – $dlINV_t$



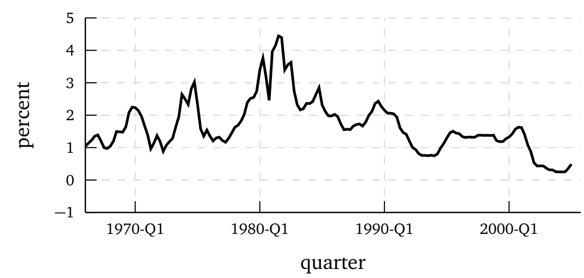
(d) Per Capita Real Wage Growth – $dlWAGE_t$



(e) Per Capita Hours Index – $lHOURS_t$



(f) Inflation – dlP_t



(g) Federal Funds Rate – $FEDFUNDS_t$

Notes: The data is adopted from the FORTRAN codes provided by Herbst (2015) and covers 1966:Q1 to 2004:Q4. The construction follows that of Smets and Wouters (2007) and is explained in detail by Herbst and Schorfheide (2016). **Source:** Herbst (2015).

may be rewritten as a rational expectations model of the form

$$B\,\mathbb{E}_t \begin{pmatrix} \mathbf{w}_{t+1}^{(m)} \\ \mathbf{y}_{t+1}^{(m)} \end{pmatrix} = A \begin{pmatrix} \mathbf{w}_{t}^{(m)} \\ \mathbf{y}_{t}^{(m)} \end{pmatrix},$$

where $\mathbf{w}_t^{(m)} = \begin{pmatrix} \mathbf{z}_t^{(m)} & \mathbf{x}_t^{(m)} \end{pmatrix}^T$. As shown by Klein (2000), this model can be solved using the generalized Schur decomposition. The resulting policy function takes the form:

$$\mathbf{x}_{t+1}^{(m)} = L_x^x\,\mathbf{x}_t^{(m)} + L_z^x\,\mathbf{z}_t^{(m)}, \tag{E.63a}$$

$$\mathbf{y}_{t}^{(m)} = L_x^y\,\mathbf{x}_t^{(m)} + L_z^y\,\mathbf{z}_t^{(m)}, \tag{E.63b}$$

$$\mathbf{z}_{t+1}^{(m)} = \boldsymbol{\eta}_{t+1}, \tag{E.63c}$$

where the vector $\boldsymbol{\eta}_t$ collects the residuals η_t^a, η_t^b, η_t^g, η_t^i, η_t^r, η_t^p, and η_t^w. Further, we denote $\tilde{L}_z^y$ and $\tilde{L}_x^y$ as the rows of L_x^y and L_z^y that correspond to the endogenous variables $\bar{y}_t$, $\bar{c}_t$, $\bar{i}_t$, $\bar{w}_t$, l_t, π_t, and r_t, so that

$$\mathbf{y}_t = \begin{pmatrix} \bar{\gamma} & \bar{\gamma} & \bar{\gamma} & \bar{\gamma} & \bar{l} & \bar{\pi} & \bar{r} \end{pmatrix}^T + \tilde{L}_x^y\,\mathbf{x}_t^{(m)} + \tilde{L}_z^y\,\mathbf{z}_t^{(m)}.$$

E.7 State-space representation

Using the policy function (E.63), we may rewrite the solved model in terms of the SSM (2) by defining $\mathbf{w}_t$, $\mathbf{v}_{t,z}$, $\mathbf{h}$, $\mathbf{H}$, $\mathbf{F}$ and $\mathbf{Q}$ as

$$\mathbf{v}_{t,z} = \begin{pmatrix} \eta_t^a & \eta_t^b & \eta_t^g & \eta_t^i & \eta_t^r & \eta_t^p & \eta_t^w \end{pmatrix}^T, \qquad \mathbf{w}_t = \mathbf{w}_t^{(m)},$$

$$\mathbf{h} = \begin{pmatrix} \bar{\gamma} & \bar{\gamma} & \bar{\gamma} & \bar{\gamma} & \bar{l} & \bar{\pi} & \bar{r} \end{pmatrix}^T, \qquad \mathbf{H}_z = \tilde{L}_z^y, \qquad \mathbf{H}_x = \tilde{L}_x^y,$$

$$\mathbf{Q}_z = \operatorname{diag}\left(\sigma_a^2, \sigma_b^2, \sigma_g^2, \sigma_i^2, \sigma_r^2, \sigma_p^2, \sigma_w^2\right), \qquad \mathbf{F}_z = \begin{pmatrix} \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad \mathbf{F}_x = \begin{pmatrix} L_z^x & L_x^x \end{pmatrix}.$$

Consequently, the model satisfies the preconditions of Proposition 3.1, provided the matrix is $\tilde{L}_x^y$ is non-singular.

E.8 Parameters and steady-state

The model has 36 parameters to be estimated. The prior distributions of these parameters are displayed in Table 6. Further, the model contains the 5 fixed parameters:

$$\delta = 0.03, \qquad \lambda_w = 1.50, \qquad g_y = 0.18, \qquad \varepsilon_p = 10.00, \qquad \varepsilon_w = 10.00.$$

as well as the 15 dependent parameters defined by

$$
\begin{aligned}
\pi^* &= \frac{\bar{\pi}}{100} + 1, \\
\gamma &= \frac{\bar{\gamma}}{100} + 1, \\
\beta &= \frac{1}{\frac{\tilde{\beta}}{100} + 1}, \\
\bar{r} &= \frac{100\,\gamma^{\sigma_c}\,\pi^*}{\beta} - 100, \\
\bar{\beta} &= \frac{\beta}{\gamma^{\sigma_c}}, \\
r_{ss}^k &= \delta + \frac{\gamma^{\sigma_c}}{\beta} - 1, \\
w_{ss} &= \frac{1}{\left(\frac{\alpha^\alpha\,(1-\alpha)^{1-\alpha}}{\Phi\,{r_{ss}^k}^\alpha}\right)^{\frac{1}{\alpha-1}}}, \\
i_{k,\gamma} &= \frac{\delta-1}{\gamma} + 1, \\
i_k &= \gamma\left(\frac{\delta-1}{\gamma} + 1\right), \\
l_k &= -\frac{r_{ss}^k\,(\alpha-1)}{\alpha\,w_{ss}}, \\
k_y &= \Phi\,{l_k}^{\alpha-1}, \\
i_y &= i_k\,k_y, \\
c_y &= 1 - i_k\,k_y - g_y, \\
z_y &= k_y\,r_{ss}^k, \\
wl_c &= -\frac{k_y\,r_{ss}^k\,(\alpha-1)}{\alpha\,c_y\,\lambda_w}.
\end{aligned}
$$