ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Final Version Master Thesis Econometrics and Management Science

Markov Switching Models, Dynamic Factor Models, or both? Comparing models in an elaborate simulation study and an application in macroeconomics.

Author: Edel, Matthijs*

Student number: 509993

Supervisor: Dr. E. Kole§

Second assessor: Dr. R-J. Lange¶

November 29, 2022

The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Abstract

Financial and macroeconomic time series, like stock prices, inflation, and GDP, are crucial time series for investors, economists, and policy-makers. A plethora of models have been examined to accurately forecast these, four of which are the Markov switching model, the dynamic factor model with constant parameters and with changing parameters, and the Markov switching dynamic factor model, of which the last two are hardly investigated. To get better insights in the comparison of the aforementioned models, this paper scrutinises the models in both a simulation study in which data is generated from multiple vector autoregressive models, and an application where the series used are annual GDP growth, inflation rate, and unemployment rate in the G7 countries. In the simulation study, the Markov switching dynamic factor model performs best for all vector autoregressive specifications. In the application, the Markov switching model is the best performing model.

*509993me@student.eur.nl

§kole@ese.eur.nl

¶lange@ese.eur.nl

Table of contents

1 Introduction **4**

2 Models **7**

2.1 Univariate Markov Switching Autoregressive Model 7

2.2 Dynamic Factor (Mixture) Model 10

2.3 Markov Switching Dynamic Factor Model 13

2.4 Model Comparison 16

3 Simulation Study **17**

3.1 Simulation Procedure 17

3.2 Results 18

4 Empirical Application **21**

4.1 Data 22

4.2 Methodology 26

4.3 Results 28

4.4 Robustness 29

5 Conclusion **30**

A Derivation of the update step of the Hamilton filter (Hamilton (1989)) **35**

B Derivation of the inner expectation of the Kim smoother (Hamilton (1989)) **36**

C Derivation of p^*_{ijt} in the E-step of the EM algorithm of the Markov switching autoregressive model (Hamilton (1990)) **37**

D Derivation of parameter expressions in the M-step of the EM algorithm of the Markov switching autoregressive model (Hamilton (1990)) **38**

E Derivation of the prediction step in the Kalman filter **39**

F Derivation of the update step in the Kalman filter **40**

G Derivation of the inner expectation in the Kalman smoother **41**

H Derivation of parameter expressions in the M-step of the EM algorithm of the dynamic factor mixture model **42**

I Derivation of $P_{t,t-1|T}$ in the E-step of the EM algorithm of the dynamic factor mixture model 44

J Elaboration on the parameters in the simulation study 45

K Scree plots of the Principal Component Analysis (PCA) 46

L Short description of the programming code 48

1 Introduction

Since financial and macroeconomic time series, such as stock prices, inflation, and GDP, are crucial time series for investors, economists, and policy-makers, it is desirable to be able to accurately forecast these. Many models have been developed to do so, of which perhaps the most popular models are the ARMA and GARCH models. In order to increase the forecast accuracy, models have been developed to examine multiple time series simultaneously, either across variables, or across countries. A popular model in this class is the vector autoregressive (VAR) model, which has been investigated by many. For example, Watson (1994) surveys developments in VAR models, and Nakajima et al. (2011) provide an overview of the estimation methods in time varying VAR models.

Notwithstanding the fact that the VAR model works well, it might not always be the best choice. Namely, in this model, the number of parameters grows very large as one uses more time series, and estimation of the model then becomes cumbersome, or sometimes even impossible. Hence, dynamic factor models have become very popular. These models have five main advantages, elaborated on in Stock and Watson (2016). They are suited to fit macroeconomic data, they are consistent with macroeconomic theory, large data sets can be estimated without computational difficulties, they are suited for tasks of a macroeconomist, and many macroeconomic shocks can be explored. Moreover, these models have been extensively discussed in the literature. For instance, Molenaar (1985) uses a dynamic factor model to investigate a polygraphic record, and Ng et al. (1992) use a dynamic factor model to examine stock returns. Also, Stock and Watson (2011) discuss several applications of dynamic factor models, and Stock and Watson (2012) compare them with a variety of other shrinkage methods.

Even though these models are extremely useful and relatively easy to estimate, they are not always the best choice. These models assume a certain relationship between the factors in the model to be the same at all times. However, when crises strike, and the economy changes rapidly, these relationships may differ, and these models may fail. Hence, we may allow the parameters to be different in recessions than in expansions. This can be done in two ways. One way is to model the state separately, and then use a VAR with observed states (Alloza (2017), Gonçalves et al. (2022)). Another way of incorporating the state of the economy is to model it as a latent variable, which is done in Markov switching models. Examples of this model can be found in Hamilton (1989), who analyses gross national product, and in Goodwin (1993) who uses the model to analyse business cycles in several economies. The disadvantage of these models is that the number of parameters in the VAR model grows even more rapidly, which makes it hard to investigate the time series simultaneously, and may lead to lower forecast accuracy.

If we combine the two aforementioned models, we obtain a dynamic factor model with different parameters in different states. Here we have the best of two worlds, both investigating multiple time series simultaneously while keeping the model parsimonious, and allowing for switching in the state of the economy. In the case where the state of the economy is modelled as a latent variable, we obtain a Markov switching dynamic factor model (MSDFM), which is a promising model. However, in this model we have multiple layers of latency, which makes estimation not straightforward anymore. A Bayesian way of estimating the model is discussed in Kim and Nelson (1998) and Sims et al. (2008). Besides, Carstensen et al. (2020) uses an elastic net to estimate the model. An approximate maximum likelihood approach is considered in Kim (1994) and Kim and Yoo (1995). One can also choose to model the state separately, and then use a dynamic factor model (DFM) with observed states, which we call the dynamic factor mixture model (DFMM) and which has been hardly investigated.

The goal of this paper is to scrutinise the two models mentioned in the previous paragraph by focusing more on the comparison with other models. As opposed to the separate estimation of dynamic factor models or Markov switching models, not much literature on the combination of the two exists yet. The dynamic factor mixture model is hardly ever investigated. Regarding the MSDFM, applications are provided in Chauvet and Piger (2008), who use it as a business cycle dating method, and in Akay et al. (2013), who use it to model hedge fund contagion and risk-adjusted returns. Furthermore, Chauvet and Potter (2013) compare the performance of the MSDFM with that of other models when making point forecasts for U.S. output growth. However, a lot can still be done when comparing the model to other models.

To do so, we conduct a simulation study and we consider a data set to compare the models. The models we compare are the univariate Markov switching autoregressive Model (UMSARM), a dynamic factor model excluding Markov switching, a dynamic factor model excluding Markov switching but with different parameters in expansions and recessions (dynamic factor mixture model), and the Markov switching dynamic factor model.

In the simulation study, we generate eight different data sets of length 1100 according to several VAR structures, each having one lag, but all having different autoregressive parameters and/or a different number of series. Furthermore, there are two different states, which are generated via a Markov process, in which the VAR has different parameters. The different autoregressive parameters and number of series are investigated to examine how the models compare to each other under different circumstances, such that a comprehensive comparison of the models can be made. For each of the eight simulations, we estimate our models based on the first 1000 observations and we make 100 one-step ahead point forecasts. We then rank our models based on their forecast performance, which is assessed by their mean squared prediction errors (MSPEs), and is compared with a Diebold Mariano (DM) test.

We find that the MSDFM outperforms the other models with a large difference. The second best performing model is the DFM, which is closely followed by the DFMM. Lastly, we find that the UMSARM performs worst. Besides, the difference in performance between the dynamic factor models and the UMSARM becomes smaller as the parameters in the VAR become smaller values. In addition, the difference in performance between the dynamic factor models and the UMSARM gets smaller as the number of series in the VAR gets larger.

In the application, we consider a data set including three macroeconomic variables, annual GDP growth, unemployment rate, and inflation rate, for a total of seven countries, Canada, France, Germany, Italy, Japan, the UK, and the USA, in the period from 1991Q1 until 2021Q4. We use eleven different subsets to compare the one-step ahead point forecast performance across our models, which will again be assessed by the MSPEs and the corresponding DM test. The subsets are created as follows. We consider, for each country separately, all variables as one multivariate time series. Furthermore, we consider, for each variable separately, all countries as one multivariate time series. Finally, we consider the whole data set, countries and variables combined, as one multivariate time series. To estimate the models, we use the observations from 1991Q1 until 2011Q4. Thereafter, one-step ahead forecasts are made until 2013Q4. Then, the models are estimated from 1991Q1 until 2013Q4, and one-step ahead forecasts are made until 2015Q4, and so forth until we obtain a forecast for 2021Q4. Besides, we check the robustness of our results by doing the analysis with the same data subsets, but with the observations during the COVID-19 pandemic omitted (2020Q1-2021Q4).

In contrast with the simulation study, we find that the UMSARM performs the best in the application. The DFM performs worst, and the DFMM and the MSDFM perform relatively well, but their parameters cannot be identified at all times. When the observations during the COVID-19 pandemic are omitted, we again find that the UMSARM performs the best. Hence, if one examines a data set and believes that the data originates from a VAR, but there are not enough observations to estimate a VAR, the MSDFM might be their best option. When examining macroeconomic variables, using the UMSARM is the best option, since the addition of dynamic factors does not yield any additional performance.

We contribute to the existing literature by defining the DFMM and deriving the expressions to estimate its parameters. Furthermore, we contribute by comparing the MSDFM to other models. This paper is the first to compare the MSDFM with other models through a simulation study, and we find that it outperforms the UMSARM, the DFM, and the DFMM when examining a VAR. Moreover, this paper is among the first to compare the MSDFM to other models in an applied setting, where we find that it outperforms the DFM when examining macroeconomic variables.

The structure of this paper is as follows. In Section 2, the models are specified, and we discuss how they are compared. Then, we elaborate on the simulation study in Section 3, where the simulation procedure is explained in 3.1 and the results are shown in 3.2. Thereafter, we discuss the application in Section 4. The data is described in 4.1, the methodology in 4.2, and the results are shown in 4.3. Moreover, a robustness check is presented in 4.4. Ultimately, in Section 5, conclusions are drawn and suggestions for further research are given.

2 Models

In this section, we elaborate on the models that we use. We start with the univariate Markov switching autoregressive Model (UMSARM), we then explain the dynamic factor model (DFM) and the dynamic factor mixture model (DFMM). The last model we discuss is the Markov switching dynamic factor model (MSDFM). For each model, we specify the model itself, how the latent variables are filtered, and in what way the model is estimated and identified. Finally, we explain how the forecasts are made. We end this section by explaining how the models are compared.

2.1 Univariate Markov Switching Autoregressive Model

The univariate Markov switching autoregressive model that we use looks as follows:

$$y_t = \boldsymbol{\beta}'_{S_t} \boldsymbol{x}_t + \varepsilon_t, \tag{1}$$

where $\varepsilon_t \sim \mathrm{N}(0, \sigma^2_{S_t})$, and $\boldsymbol{x}_t = \begin{bmatrix} 1, & y_{t-1} \end{bmatrix}'$, with y_t the dependent variable for $t = 1, \dots, T$. Here, S_t is the state at time t which is an unobserved Markov process described by $\mathrm{P}[S_t = i | S_{t-1} = j] = p_{ij}$, with $i, j = 1, \dots, m$, with m being the number of states. Lastly, because of the aforementioned rapid increase in parameters and the corresponding problems if we make this model multivariate, we decided to keep this model univariate.

To estimate the current state, S_t, in the Markov switching autoregressive model, the Hamilton filter is used, together with the Kim smoother. The Hamilton filter consists of two steps, a prediction step and an update step. The prediction step looks as follows. Suppose that we know the parameters in the model. Then, let $\boldsymbol{\xi}_t$ be a vector with a 1 on the true state at time t and a 0 on the other states. Now we can write our estimate for the states, based on $\mathcal{I}_t$, as $\hat{\boldsymbol{\xi}}_{t|t}$.

Our prediction step can then be written as follows:

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \boldsymbol{P}\hat{\boldsymbol{\xi}}_{t-1|t-1}, \tag{2}$$

with $\boldsymbol{P}$ the transition matrix with $P_{ij} = \mathrm{P}[S_t = i|S_{t-1} = j] = p_{ij}$. Besides, we put random numbers in $\hat{\boldsymbol{\xi}}_{1|1}$, ensuring that they add up to one. We now update our belief based on observation y_t as follows[1]:

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{\begin{bmatrix} f(\boldsymbol{y}_t|S_t=1) \\ \vdots \\ f(\boldsymbol{y}_t|S_t=m) \end{bmatrix} \odot \hat{\boldsymbol{\xi}}_{t|t-1}}{\begin{bmatrix} 1 \cdots 1 \end{bmatrix} \left(\begin{bmatrix} f(\boldsymbol{y}_t|S_t=1) \\ \vdots \\ f(\boldsymbol{y}_t|S_t=m) \end{bmatrix} \odot \hat{\boldsymbol{\xi}}_{t|t-1} \right)}. \tag{3}$$

Once we have all observations, we can iterate back and smooth our estimates for all $\boldsymbol{\xi}_t$ based on the full sample size ($\mathcal{I}_T$). To do so, we make use of the law of iterated expectations[2] and we write $\hat{\boldsymbol{\xi}}_{t|T} = \mathbb{E}[\xi_t|\mathcal{I}_T] = \mathbb{E}[\mathbb{E}[\xi_t|\xi_{t+1}, \mathcal{I}_T]|\mathcal{I}_T]$. For the inner expectation holds that $\mathbb{E}[\xi_t|\xi_{t+1}, \mathcal{I}_T] = \hat{\boldsymbol{\xi}}_{t|t} \odot \boldsymbol{P}'[\boldsymbol{\xi}_{t+1} \oslash \hat{\boldsymbol{\xi}}_{t+1|t}]$[3]. Now $\hat{\boldsymbol{\xi}}_{t|T} = \mathbb{E}[\hat{\boldsymbol{\xi}}_{t|t} \odot \boldsymbol{P}'[\boldsymbol{\xi}_{t+1} \oslash \hat{\boldsymbol{\xi}}_{t+1|t}]|\mathcal{I}_T]$, and thus we can write the smoothing as follows:

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \boldsymbol{P}'[\hat{\boldsymbol{\xi}}_{t+1|T} \oslash \hat{\boldsymbol{\xi}}_{t+1|t}]. \tag{4}$$

The parameters in the model are estimated by the Expectation Maximisation (EM) algorithm. In this algorithm, we augment the likelihood function by S_t to obtain the so-called complete data likelihood function. This complete data likelihood can be written as follows:

$$f(y_{2:T}, S_{1:T}|\mathcal{I}_T, \boldsymbol{\theta}, \boldsymbol{P}, \boldsymbol{\rho}) = \prod_{t=2}^{T} \left[\prod_{i,j=1}^{n} \Big(f_i(y_t|\boldsymbol{\theta_i}) p_{ij} \Big)^{\delta_{ijt}} \right] \left\{ \prod_{j=1}^{n} \rho_j^{\delta_{j1}} \right\}, \tag{5}$$

with the Kronecker delta, δ_{ijt}, a random variable being 1 if $S_t = i$ and $S_{t-1} = j$, and 0 otherwise, and δ_{jt} being a random variable equalling 1 if $S_t = j$ and 0 otherwise. Furthermore, ρ_j is the probability that $S_1 = j$, and $f_i(y_t) = \phi(y_t; \beta_i y_{t-1}, \sigma_i^2)$ is the density of y_t if $S_t = i$. The log likelihood can then be written as follows:

$$\log\, f(y_{2:T}, S_{1:T}|\mathcal{I}_T, \boldsymbol{\theta}, \boldsymbol{P}, \boldsymbol{\rho}) = \sum_{t=2}^{T} \left[\sum_{i,j=1}^{n} \delta_{ijt} \log\big(f_i(y_t|\boldsymbol{\theta_i}) p_{ij} \big) \right] + \sum_{j=1}^{n} \delta_{j1} \log\big(\rho_j\big). \tag{6}$$

[1]The derivation can be found in Appendix A, which is from Hamilton (1989).
[2]$\mathbb{E}[Y] = \mathbb{E}_X[\mathbb{E}_Y[Y|X]]$.
[3]The derivation can be found in Appendix B, which is from Hamilton (1989).

Now the expected log likelihood is the following:

$$\mathbb{E}[\log f(y_{2:T}, S_{1:T}|\mathcal{I}_T, \boldsymbol{\theta}, \boldsymbol{P}, \boldsymbol{\rho})] = \sum_{t=2}^{T} \left[\sum_{i,j=1}^{n} p^*_{ijt} \log\big(f_i(y_t|\boldsymbol{\theta_i})p_{ij}\big) \right] + \sum_{j=1}^{n} p^*_{j1} \log\big(\rho_j\big), \tag{7}$$

with $p^*_{ijt} = \mathrm{P}[S_t = i, S_{t-1} = j|\mathcal{I}_T]$, and $p^*_{jt} = \mathrm{P}[S_t = j|\mathcal{I}_T]$. The latter is part of the output of the Kim smoother, the former we must add to the Kim smoother and reads as follows[4]:

$$p^*_{ijt} = \left[\boldsymbol{P} \odot (\hat{\boldsymbol{\xi}}_{t|T} \hat{\boldsymbol{\xi}}'_{t-1|t-1}) \oslash (\hat{\boldsymbol{\xi}}_{t|t-1}[1 \cdots 1])\right]_{ij}. \tag{8}$$

Once we have the expected log likelihood, we maximise it with respect to the parameters, and we get the following expressions[5]:

$$\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\xi}}_{1|T}, \tag{9}$$

$$\hat{p}_{kl} = \frac{\sum_{t=2}^{T} p^*_{klt}}{\sum_{t=2}^{T} p^*_{l(t-1)})}, \tag{10}$$

$$\hat{\boldsymbol{\beta}}_k = (\boldsymbol{X}'_l P^*_k \boldsymbol{X}_l)^{-1} \boldsymbol{X}'_l P^*_k \boldsymbol{y}, \tag{11}$$

$$\hat{\sigma}^2_k = \frac{\sum_{t=2}^{T} p^*_{kt}(y_t - \hat{\boldsymbol{\beta}}'_k \boldsymbol{x}_t)^2}{\sum_{t=2}^{T} p^*_{kt}}. \tag{12}$$

Here, P^*_k is a $(T-1) \times (T-1)$ diagonal matrix with $p^*_{k(t+1)}$ on the t^{th} position and $\boldsymbol{y}$ is a vector of length $T-1$ with y_{t+1} on the t^{th} position. Furthermore, $\boldsymbol{X}_l$ is $(T-1) \times 2$ matrix with $\boldsymbol{x}'_t$ on the t^{th} row. We initialise the algorithm with random numbers for all parameters. We then run the Hamilton filter and Kim smoother and estimate the new parameters. We keep repeating this until the parameters converge or until we have reached 1000 iterations.

Once the parameters are estimated, the one-step ahead forecasts in the univariate Markov switching model are made as follows:

$$\begin{aligned}
\hat{y}_{T+1|T} &= \mathbb{E}\left[\sum_{k=1}^{m} \mathrm{I}[S_{T+1} = k]\Big(\boldsymbol{\beta}'_k \boldsymbol{x}_{T+1}\Big) + \varepsilon_{T+1}|\mathcal{I}_T\right] \\
&= \sum_{k=1}^{m} \mathrm{P}[S_{T+1} = k]\Big(\hat{\boldsymbol{\beta}}'_k \boldsymbol{x}_{T+1}\Big) \\
&= \hat{\boldsymbol{\xi}}'_{T+1|T} \begin{bmatrix} \hat{\boldsymbol{\beta}}'_1 \boldsymbol{x}_{T+1} \\ \vdots \\ \hat{\boldsymbol{\beta}}'_m \boldsymbol{x}_{T+1} \end{bmatrix}.
\end{aligned} \tag{13}$$

[4]The derivation can be found in Appendix C, which is from Hamilton (1990).
[5]The derivation can be found in Appendix D, which is from Hamilton (1990).

2.2 Dynamic Factor (Mixture) Model

We use two dynamic factor models, one with constant parameters, and one with different parameters in expansions and recessions, which we call the mixture model, looking as follows:

$$\boldsymbol{y}_t = \boldsymbol{c}_{S_t} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \tag{14}$$

$$\boldsymbol{f}_t = \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t, \tag{15}$$

where $\boldsymbol{y}_t$ has length n for $t = 1, \ldots, T$, $\boldsymbol{f}_t$ are latent factors, and $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{R}_{S_t})$ and $\boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_{S_t})$. Here, S_t is the state at time t which is 0 when there is an expansion and 1 when there is a recession. The dynamic factor model with constant parameters has the same specification except that the subscript S_t is omitted from that model because the parameters are the same in both expansions and recessions. The current state S_t is given during estimation, and is modelled separately for forecasting and is then assumed to be equal to S_{t-1}.

The latent factors $\boldsymbol{f}_t$ are estimated by the Kalman filter and smoother. Henceforth, everything in this section works the same for the dynamic factor model with constant parameters, except that any superscripts or subscripts relating to states are omitted since there is only one state in the dynamic factor model with constant parameters.

Like the Hamilton filter, the Kalman filter consists of a prediction step and an update step. The prediction step looks as follows. Suppose that we know the parameters in the model. Then, let $\hat{\boldsymbol{f}}_{t|t}$ be our estimate for $\boldsymbol{f}_t$, based on $\mathcal{I}_t$, and let $\boldsymbol{P}_{t|t}$ be the uncertainty of that estimate. Now we can write the prediction step as follows[6]:

$$\hat{\boldsymbol{f}}^k_{t|t-1} = \boldsymbol{A}_k\hat{\boldsymbol{f}}_{t-1|t-1}, \tag{16}$$

$$\boldsymbol{P}^k_{t|t-1} = \boldsymbol{A}_k\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}'_k + \boldsymbol{Q}_k, \tag{17}$$

with $\mathbf{0}$, the unconditional mean of $\boldsymbol{f}_t$, as $\hat{\boldsymbol{f}}_{0|0}$ and a diagonal matrix with 10^6 on the diagonal elements as $\boldsymbol{P}_{0|0}$. Here, $\hat{\boldsymbol{f}}^k_{t|t-1}$ is our estimate for $\boldsymbol{f}_t$, based on $\mathcal{I}_{t-1}$, and assuming that the process is in state k at time t. Furthermore, $\boldsymbol{P}^k_{t|t}$ is the uncertainty of that estimate.

Now we can write the update step, in which we update our estimate for $\boldsymbol{f}_t$ and the corresponding uncertainty $\boldsymbol{P}_t$, based on observation $\boldsymbol{y}_t$, and the observation of S_t. Since S_t is now known, we only update the factors and the corresponding uncertainties of Equations 16 and 17 in the correct state. These updates are now simply denoted as $\hat{\boldsymbol{f}}_{t|t}$ and $\boldsymbol{P}_{t|t}$, and read as follows[7]:

[6] The derivation can be found in Appendix E.
[7] The derivation can be found in Appendix F.

$$\hat{\boldsymbol{f}}_{t|t} = \hat{\boldsymbol{f}}_{t|t-1}^{S_t} + \boldsymbol{P}_{t|t-1}^{S_t}\boldsymbol{\Lambda}_{S_t}'\left(\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}_{t|t-1}^{S_t}\boldsymbol{\Lambda}_{S_t}' + \boldsymbol{R}_{S_t}\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\hat{\boldsymbol{f}}_{t|t-1}^{S_t}\right), \tag{18}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1}^{S_t} - \boldsymbol{P}_{t|t-1}^{S_t}\boldsymbol{\Lambda}_{S_t}'\left(\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}_{t|t-1}^{S_t}\boldsymbol{\Lambda}_{S_t}' + \boldsymbol{R}_{S_t}\right)^{-1}\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}_{t|t-1}^{S_t}. \tag{19}$$

Once we have all observations, we can iterate back and smooth our estimates for the factors based on the full sample size ($\mathcal{I}_T$). Hence, we are interested in $\hat{\boldsymbol{f}}_{t|T} = \mathbb{E}[\boldsymbol{f}_t|\mathcal{I}_T] = \mathbb{E}[\mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, \mathcal{I}_T]|\mathcal{I}_T]$, where we have again made use of the law of iterated expectations. We have that $\mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, \mathcal{I}_T] = \mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, S_{t+1}, \mathcal{I}_t]$ because all future information that is relevant for $\boldsymbol{f}_t$ is included in $\boldsymbol{f}_{t+1}$ and S_{t+1}. Since $\mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, S_{t+1}, \mathcal{I}_t] = \hat{\boldsymbol{f}}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{A}_{S_{t+1}}'(\boldsymbol{P}_{t+1|t}^{S_t})^{-1}(\boldsymbol{f}_{t+1} - \hat{\boldsymbol{f}}_{t+1|t}^{S_t})$[8], we can write the following equation for the smoothed factor:

$$\begin{aligned}\hat{\boldsymbol{f}}_{t|T} &= \mathbb{E}[\hat{\boldsymbol{f}}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{A}_{S_{t+1}}'\boldsymbol{P}_{t+1|t}^{-1}(\boldsymbol{f}_{t+1} - \hat{\boldsymbol{f}}_{t+1|t})|\mathcal{I}_T] \\ &= \hat{\boldsymbol{f}}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{A}_{S_{t+1}}'(\boldsymbol{P}_{t+1|t}^{S_t})^{-1}(\hat{\boldsymbol{f}}_{t+1|T} - \hat{\boldsymbol{f}}_{t+1|t}^{S_t})\end{aligned} \tag{20}$$

Furthermore, to find the smoothed variance of $\boldsymbol{f}_t$, $\boldsymbol{P}_{t|T}$, we make use of the law of total variance[9] and we write $\boldsymbol{P}_{t|T} = \mathrm{Var}[\boldsymbol{f}_t|\mathcal{I}_T] = \mathrm{Var}[\mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, \mathcal{I}_T]|\mathcal{I}_T] + \mathbb{E}[\mathrm{Var}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, \mathcal{I}_T]|\mathcal{I}_T]$. Hence, we can write

$$\boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t} - \boldsymbol{P}_{t|t}\boldsymbol{A}_{S_{t+1}}'(\boldsymbol{P}_{t+1|t}^{S_t})^{-1}\big(\boldsymbol{P}_{t+1|t}^{S_t} - \boldsymbol{P}_{t+1|T}\big)(\boldsymbol{P}_{t+1|t}^{S_t})^{-1}\boldsymbol{A}_{S_{t+1}}\boldsymbol{P}_{t|t}. \tag{21}$$

As in the univariate Markov switching model, the parameters in the dynamic factor models are estimated by the EM algorithm. Here we augment the likelihood function by f_t to obtain the so-called complete data likelihood function. This complete data likelihood can be written as follows:

$$\begin{aligned}\log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta}) \propto\ & \frac{1}{2}\sum_{t=1}^{T}\log|\boldsymbol{R}_{S_t}^{-1}| - \frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t)'\boldsymbol{R}_{S_t}^{-1}(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t) \\ &+ \frac{1}{2}\sum_{t=2}^{T}\log|\boldsymbol{Q}_{S_t}^{-1}| - \frac{1}{2}\sum_{t=2}^{T}(\boldsymbol{f}_t - \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1})'\boldsymbol{Q}_{S_t}^{-1}(\boldsymbol{f}_t - \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1}),\end{aligned} \tag{22}$$

which leads to the following parameter estimates[10]:

[8] The derivation can be found in Appendix G.

[9] $\mathrm{Var}[Y] = \mathrm{Var}[\mathbb{E}[Y|X]] + \mathbb{E}[\mathrm{Var}[Y|X]]$.

[10] The derivation can be found in Appendix H.

$$\hat{\boldsymbol{A}}_k = \Big(\sum_{t=2}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t-1|T} + \boldsymbol{P}_{t,t-1|T}) \Big) \times \Big(\sum_{t=2}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}_{t-1|T} + \boldsymbol{P}'_{t-1|T}) \Big)^{-1}, \tag{23}$$

$$\hat{\boldsymbol{\Lambda}}_k = \Big(\sum_{t=1}^{T} \mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)\hat{\boldsymbol{f}}'_{t|T} \Big) \Big(\sum_{t=1}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T}) \Big)^{-1}, \tag{24}$$

$$\hat{\boldsymbol{c}}_k = \frac{\sum_{t=1}^{T} \mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{\Lambda}}_k \hat{\boldsymbol{f}}_{t|T})}{\sum_{t=1}^{T} \mathrm{I}[S_t = k]}, \tag{25}$$

$$\begin{aligned} \hat{\boldsymbol{R}}_k = \frac{1}{\sum_{t=1}^{T} \mathrm{I}[S_t = k]} \sum_{t=1}^{T} \mathrm{I}[S_t = k] \Big(&(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)' - \hat{\boldsymbol{\Lambda}}_k \hat{\boldsymbol{f}}_{t|T}(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)' \\ &- (\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)\hat{\boldsymbol{f}}'_{t|T}\hat{\boldsymbol{\Lambda}}'_k + \hat{\boldsymbol{\Lambda}}_k(\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T})\hat{\boldsymbol{\Lambda}}'_k \Big), \end{aligned} \tag{26}$$

$$\begin{aligned} \hat{\boldsymbol{Q}}_k = \frac{1}{\sum_{t=2}^{T} \mathrm{I}[S_t = k]} \sum_{t=2}^{T} \mathrm{I}[S_t = k] \Big(&\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T} - \hat{\boldsymbol{A}}_k(\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}'_{t,t-1|T}) \\ &- (\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t-1|T} + \boldsymbol{P}_{t,t-1|T})\hat{\boldsymbol{A}}'_k + \hat{\boldsymbol{A}}_k(\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}'_{t-1|T} + \boldsymbol{P}_{t-1|T})\hat{\boldsymbol{A}}'_k \Big), \end{aligned} \tag{27}$$

where we use, for each state k, the $\hat{\boldsymbol{c}}_k$ from the previous iteration when calculating $\hat{\boldsymbol{\Lambda}}_k$. Furthermore, we first calculate $\hat{\boldsymbol{A}}_k$, $\hat{\boldsymbol{\Lambda}}_k$, and $\hat{\boldsymbol{c}}_k$, and we plug in these estimates when calculating $\hat{\boldsymbol{R}}_k$ and $\hat{\boldsymbol{Q}}_k$.

We initialise the algorithm with random numbers for all parameters. We then run the Kalman filter and smoother and estimate the new parameters. We keep repeating this until the parameters converge or until we have reached 1000 iterations. Furthermore, to identify the model, we set the upper part of each $\hat{\boldsymbol{\Lambda}}_k$ to an $r \times r$ identity matrix, with r being the number of factors in the model, where we follow Bai and Wang (2015), who examine a DFM and set the upper part of $\hat{\boldsymbol{\Lambda}}$ to an $r \times r$ identity matrix.

To make one-step ahead forecasts, we must first forecast the state and the factors. Thereafter, we can produce the actual forecast. The state is forecast as follows: $\hat{S}_{t+1} = S_t$. Now we can write the following forecast for the factors:

$$\begin{aligned} \hat{\boldsymbol{f}}_{T+1|T} &= \mathbb{E}[\boldsymbol{A}_{S_{T+1}}\boldsymbol{f}_T + \boldsymbol{\eta}_{T+1}|\mathcal{I}_T] \\ &= \hat{\boldsymbol{A}}_{\hat{S}_{T+1}}\hat{\boldsymbol{f}}_{T|T} \\ &= \hat{\boldsymbol{A}}_{S_T}\hat{\boldsymbol{f}}_{T|T}. \end{aligned} \tag{28}$$

With this forecast for the factors, we can write the actual forecast, which looks as follows:

$$\begin{aligned}\hat{\boldsymbol{y}}_{T+1|T} &= \mathbb{E}[\boldsymbol{c}_{S_{T+1}} + \boldsymbol{\Lambda}_{S_{T+1}}\boldsymbol{f}_{T+1} + \boldsymbol{\varepsilon}_{T+1}] \\ &= \hat{\boldsymbol{c}}_{\hat{S}_{T+1}} + \hat{\boldsymbol{\Lambda}}_{\hat{S}_{T+1}}\hat{\boldsymbol{f}}_{T+1|T} \\ &= \hat{\boldsymbol{c}}_{S_T} + \hat{\boldsymbol{\Lambda}}_{S_T}\hat{\boldsymbol{f}}_{T+1|T}.\end{aligned} \tag{29}$$

2.3 Markov Switching Dynamic Factor Model

For the Markov switching dynamic factor model we use the following specification, inspired by Brownlees and Kole (2022) and Kim (1994):

$$\boldsymbol{y}_t = \boldsymbol{c}_{S_t} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \tag{30}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\mathcal{B}}_{S_t}\boldsymbol{\mu}_{t-1} + \begin{bmatrix}\boldsymbol{\eta}_t \\ \mathbf{0}\end{bmatrix}, \tag{31}$$

where $\boldsymbol{y}_t$ has length n for $t = 1, \ldots, T$, $\boldsymbol{f}_t$ are latent factors, and $\boldsymbol{\varepsilon}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{R}_{S_t})$ and $\boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{Q}_{S_t})$. Just as in Equation 1, S_t is the state at time t which is an unobserved Markov process described by $\mathrm{P}[S_t = i|S_{t-1} = j] = p_{ij}$, with $i, j = 1, \ldots, m$, with m being the number of states. Furthermore, $\boldsymbol{\mu}_t = \begin{bmatrix}\boldsymbol{f}_t \\ \boldsymbol{f}_{t-1}\end{bmatrix}$ and $\boldsymbol{\mathcal{B}}_{S_t} = \begin{bmatrix}\boldsymbol{A}_{S_t} & \mathbf{O} \\ \mathbf{I} & \mathbf{O}\end{bmatrix}$, with $\mathbf{I}$ being an $r \times r$ identity matrix, and $\mathbf{O}$ being an $r \times r$ zero matrix, with r being the number of factors in the model.

In this model, we have two layers of latency, since we have both latent factors and a latent Markov process determining the state at time t. To filter these latent variables and to estimate the model, we use the Kim filter and smoother as described Brownlees and Kole (2022). We again have a prediction step and an update step. The prediction and update steps look as follows:

$$\hat{\boldsymbol{\mu}}^{(j,k)}_{t|t-1} = \boldsymbol{\mathcal{B}}_k\hat{\boldsymbol{\mu}}^{j}_{t-1|t-1}, \tag{32}$$

$$\boldsymbol{\Sigma}^{(j,k)}_{t|t-1} = \boldsymbol{\mathcal{B}}_k\boldsymbol{\Sigma}^{j}_{t-1|t-1}\boldsymbol{\mathcal{B}}_k' + \boldsymbol{Z}\boldsymbol{Q}_k\boldsymbol{Z}', \tag{33}$$

$$\hat{\boldsymbol{\mu}}^{(j,k)}_{t|t} = \hat{\boldsymbol{\mu}}^{(j,k)}_{t|t-1} + \boldsymbol{\Sigma}^{(j,k)}_{t|t-1}\boldsymbol{Z}\boldsymbol{\Lambda}_k'\left(\boldsymbol{\Lambda}_k\boldsymbol{Z}'\boldsymbol{\Sigma}^{(j,k)}_{t|t-1}\boldsymbol{Z}\boldsymbol{\Lambda}_k' + \boldsymbol{R}_k\right)^{-1}\left(\boldsymbol{y}_t - \boldsymbol{c}_k - \boldsymbol{\Lambda}_k\boldsymbol{Z}'\hat{\boldsymbol{\mu}}^{(j,k)}_{t|t-1}\right), \tag{34}$$

$$\boldsymbol{\Sigma}^{(j,k)}_{t|t} = \boldsymbol{\Sigma}^{(j,k)}_{t|t-1} - \boldsymbol{\Sigma}^{(j,k)}_{t|t-1}\boldsymbol{Z}\boldsymbol{\Lambda}_k'\left(\boldsymbol{\Lambda}_k\boldsymbol{Z}'\boldsymbol{\Sigma}^{(j,k)}_{t|t-1}\boldsymbol{Z}\boldsymbol{\Lambda}_k' + \boldsymbol{R}_k\right)^{-1}\boldsymbol{\Lambda}_k\boldsymbol{Z}'\boldsymbol{\Sigma}^{(j,k)}_{t|t-1}. \tag{35}$$

Here, $\hat{\boldsymbol{\mu}}^{(j,k)}_{t|t-1}$ is our estimate for $\boldsymbol{\mu}_t$, given $\mathcal{I}_{t-1}$, and that $S_t = k$ and $S_{t-1} = j$. Furthermore, $\boldsymbol{\Sigma}^{(j,k)}_{t|t-1}$ is the uncertainty of $\hat{\boldsymbol{\mu}}^{(j,k)}_{t|t-1}$. Moreover, $\boldsymbol{Z} = \begin{bmatrix}\mathbf{I} \\ \mathbf{O}\end{bmatrix}$, with again $\mathbf{I}$ being an $r \times r$ identity matrix, and $\mathbf{O}$ being an $r \times r$ zero matrix.

Now we collapse the terms and we write the following:

$$\hat{\boldsymbol{\mu}}_{t|t}^{k} = \frac{\sum_{j=1}^{m} \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_t]\hat{\boldsymbol{\mu}}_{t|t}^{(j,k)}}{\mathrm{P}[S_t = k|\mathcal{I}_t]}, \tag{36}$$

$$\boldsymbol{\Sigma}_{t|t}^{k} = \frac{\sum_{j=1}^{m} \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_t]\big(\boldsymbol{\Sigma}_{t|t}^{(j,k)} + (\hat{\boldsymbol{\mu}}_{t|t}^{k} - \hat{\boldsymbol{\mu}}_{t|t}^{(j,k)})(\hat{\boldsymbol{\mu}}_{t|t}^{k} - \hat{\boldsymbol{\mu}}_{t|t}^{(j,k)})'\big)}{\mathrm{P}[S_t = k|\mathcal{I}_t]}. \tag{37}$$

Here, $\hat{\boldsymbol{\mu}}_{t|t}^{k}$ is our estimate for $\boldsymbol{\mu}_t$, given $\mathcal{I}_t$, and that $S_t = k$, with $\boldsymbol{\Sigma}_{t|t}^{k}$ the uncertainty of that estimate. To complete the filtering, we need to calculate the utilised probability terms, which we do as follows:

$$f(\boldsymbol{y}_t, S_{t-1} = j, S_t = k|\mathcal{I}_{t-1}) = f(\boldsymbol{y}_t|S_{t-1} = j, S_t = k, \mathcal{I}_{t-1}) \\ \times \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_{t-1}], \tag{38}$$

where $f(\boldsymbol{y}_t|S_{t-1} = j, S_t = k, \mathcal{I}_{t-1}) = \phi(\boldsymbol{y}_t - \boldsymbol{c}_k - \boldsymbol{\Lambda}_k \boldsymbol{Z}' \hat{\boldsymbol{\mu}}_{t|t-1}^{(j,k)}; \boldsymbol{0}, \boldsymbol{\Lambda}_k \boldsymbol{Z}' \boldsymbol{\Sigma}_{t|t-1}^{(j,k)} \boldsymbol{Z} \boldsymbol{\Lambda}_k' + \boldsymbol{R}_k)$,

with ϕ the pdf of the normal distribution.

$$\mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_t] = \frac{f(\boldsymbol{y}_t, S_{t-1} = j, S_t = k|\mathcal{I}_{t-1})}{f(\boldsymbol{y}_t|\mathcal{I}_{t-1})}, \tag{39}$$

where $f(\boldsymbol{y}_t|\mathcal{I}_{t-1}) = \sum_{k=1}^{m}\sum_{j=1}^{m} f(\boldsymbol{y}_t, S_{t-1} = j, S_t = k|\mathcal{I}_{t-1})$.

$$\mathrm{P}[S_t = k|\mathcal{I}_t] = \sum_{j=1}^{m} \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_t], \tag{40}$$

$$\mathrm{P}[S_t = k, S_{t+1} = i|\mathcal{I}_t] = \mathrm{P}[S_{t+1} = i|S_t = k]\mathrm{P}[S_t = k|\mathcal{I}_t]. \tag{41}$$

To start the filtering, we set the following input values. For all k, $\hat{\boldsymbol{\mu}}_{0|0}^{k} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_{0|0}^{k} = \begin{bmatrix} 10^6\mathbf{I} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{bmatrix}$, with again $\mathbf{I}$ being an $r \times r$ identity matrix, and $\mathbf{O}$ being an $r \times r$ zero matrix. Furthermore, we set, for all j, and for all k, $\mathrm{P}[S_0 = j, S_1 = k|\mathcal{I}_0] = 1/m^2$.

Once we have all observations, we can iterate back and smooth our estimates based on the full sample size ($\mathcal{I}_T$). For this we write the following:

$$\mathrm{P}[S_{T-1} = j|\mathcal{I}_T] = \sum_{k=1}^{m} \mathrm{P}[S_{T-1} = j, S_T = k|\mathcal{I}_T], \tag{42}$$

$$\mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_T] = \frac{\mathrm{P}[S_t = k|\mathcal{I}_T] \times \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_{t-1}]}{\mathrm{P}[S_t = k|\mathcal{I}_{t-1}]}, \tag{43}$$

$$\mathrm{P}[S_{t-1} = j|\mathcal{I}_T] = \sum_{k=1}^{m} \mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_T]. \tag{44}$$

Now we can write the following:

$$\hat{\boldsymbol{\mu}}_{t|T}^{(j,k)} = \hat{\boldsymbol{\mu}}_{t|t}^{j} + \tilde{\boldsymbol{\Sigma}}_t^{(j,k)}(\hat{\boldsymbol{\mu}}_{t+1|T}^{k} - \hat{\boldsymbol{\mu}}_{t+1|t}^{(j,k)}), \tag{45}$$

$$\boldsymbol{\Sigma}_{t|T}^{(j,k)} = \boldsymbol{\Sigma}_{t|t}^{j} + \tilde{\boldsymbol{\Sigma}}_t^{(j,k)}(\boldsymbol{\Sigma}_{t+1|T}^{k} - \boldsymbol{\Sigma}_{t+1|t}^{(j,k)})\tilde{\boldsymbol{\Sigma}}_t^{(j,k)\prime}, \tag{46}$$

$$\boldsymbol{\Sigma}_{t+1,t|T}^{(j,k)} = \boldsymbol{\Sigma}_{t+1|T}^{k}\tilde{\boldsymbol{\Sigma}}_t^{(j,k)\prime}. \tag{47}$$

Here, $\tilde{\boldsymbol{\Sigma}}_t^{(j,k)} = \boldsymbol{\Sigma}_{t|t}^{j}\boldsymbol{\mathcal{B}}_k'(\boldsymbol{\Sigma}_{t+1|t}^{(j,k)})^{-1}$, $\hat{\boldsymbol{\mu}}_{t|T}^{(j,k)}$ is our estimate for $\boldsymbol{\mu}_t$, given $\mathcal{I}_T$, and that $S_{t+1} = k$ and $S_t = j$, with $\boldsymbol{\Sigma}_{t|T}^{(j,k)}$ the uncertainty of that estimate. Furthermore, $\boldsymbol{\Sigma}_{t+1,t|T}^{(j,k)}$ is the covariance of $\boldsymbol{\mu}_{t+1}$ and $\boldsymbol{\mu}_t$.

Now we can again collapse some terms and we get the following:

$$\hat{\boldsymbol{\mu}}_{t|T}^{j} = \frac{\sum_{k=1}^{m} \mathrm{P}[S_t = j, S_{t+1} = k|\mathcal{I}_T]\hat{\boldsymbol{\mu}}_{t|T}^{(j,k)}}{\mathrm{P}[S_t = j|\mathcal{I}_T]}, \tag{48}$$

$$\boldsymbol{\Sigma}_{t|T}^{j} = \frac{\sum_{k=1}^{m} \mathrm{P}[S_t = j, S_{t+1} = k|\mathcal{I}_T]\big(\boldsymbol{\Sigma}_{t|T}^{(j,k)} + (\hat{\boldsymbol{\mu}}_{t|T}^{j} - \hat{\boldsymbol{\mu}}_{t|T}^{(j,k)})(\hat{\boldsymbol{\mu}}_{t|T}^{j} - \hat{\boldsymbol{\mu}}_{t|T}^{(j,k)})'\big)}{\mathrm{P}[S_t = j|\mathcal{I}_T]}. \tag{49}$$

Again, the parameters are estimated by the EM algorithm. This EM algorithm is the same as the one for the dynamic factor mixture model, but with a few details changed since the state is not given but estimated through the Kim filter and the model is written in companion form. Hence, for the M-step in the Markov switching dynamic factor model, we have a slightly modified version of Equations 23-27, that looks as follows:

$$\begin{aligned}\begin{bmatrix}\hat{\boldsymbol{c}}_k & \hat{\boldsymbol{\Lambda}}_k\end{bmatrix} &= \Big(\sum_{t=1}^{T}\mathrm{P}[S_t = k|\mathcal{I}_T]\boldsymbol{y}_t\begin{bmatrix}1 & \hat{\boldsymbol{f}}_{t|T}^{k}{}'\end{bmatrix}\Big) \\ &\quad\times\Big(\sum_{t=1}^{T}\mathrm{P}[S_t = k|\mathcal{I}_T]\begin{bmatrix}1 & \hat{\boldsymbol{f}}_{t|T}^{k}{}' \\ \hat{\boldsymbol{f}}_{t|T}^{k}{}' & (\hat{\boldsymbol{f}}_{t|T}^{k}\hat{\boldsymbol{f}}_{t|T}^{k}{}' + \boldsymbol{Z}'\boldsymbol{\Sigma}_{t|T}^{k}\boldsymbol{Z})^{-1}\end{bmatrix}\Big),\end{aligned} \tag{50}$$

$$\begin{aligned}\hat{\boldsymbol{A}}_k &= \Big(\boldsymbol{Z}'\sum_{t=2}^{T}\sum_{j=1}^{m}\mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_T](\hat{\boldsymbol{\mu}}_{t|T}^{k}\hat{\boldsymbol{\mu}}_{t-1|T}^{(j,k)}{}' + \boldsymbol{\Sigma}_{t,t-1|T}^{(j,k)})\boldsymbol{Z}\Big) \\ &\quad\times\Big(\boldsymbol{Z}'\sum_{t=2}^{T}\sum_{j=1}^{m}\mathrm{P}[S_{t-1} = j, S_t = k|\mathcal{I}_T](\hat{\boldsymbol{\mu}}_{t-1|T}^{(j,k)}\hat{\boldsymbol{\mu}}_{t-1|T}^{(j,k)}{}' + \boldsymbol{\Sigma}_{t-1|T}^{(j,k)})\boldsymbol{Z}\Big)^{-1},\end{aligned} \tag{51}$$

$$\begin{aligned}\hat{\boldsymbol{R}}_k &= \frac{1}{\sum_{t=1}^{T}\mathrm{P}[S_t = k|\mathcal{I}_T]}\sum_{t=1}^{T}\mathrm{P}[S_t = k|\mathcal{I}_T] \\ &\quad\times\Big((\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k - \hat{\boldsymbol{\Lambda}}_k\hat{\boldsymbol{f}}_{t|T}^{k})(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k - \hat{\boldsymbol{\Lambda}}_k\hat{\boldsymbol{f}}_{t|T}^{k})' + \hat{\boldsymbol{\Lambda}}_k(\boldsymbol{Z}'\boldsymbol{\Sigma}_{t|T}^{k}\boldsymbol{Z})\hat{\boldsymbol{\Lambda}}_k'\Big),\end{aligned} \tag{52}$$

$$
\begin{aligned}
\hat{\boldsymbol{Q}}_k = & \frac{1}{\sum_{t=2}^{T} \mathrm{P}[S_t = k | \mathcal{I}_T]} \\
& \times \sum_{t=2}^{T} \sum_{j=1}^{m} \mathrm{P}[S_{t-1} = j, S_t = k | \mathcal{I}_T] \\
& \times \Big(\boldsymbol{Z}' (\hat{\boldsymbol{\mu}}_{t|T}^{k} - \hat{\boldsymbol{\mathcal{B}}}_k \hat{\boldsymbol{\mu}}_{t-1|T}^{(j,k)}) (\hat{\boldsymbol{\mu}}_{t|T}^{k} - \hat{\boldsymbol{\mathcal{B}}}_k \hat{\boldsymbol{\mu}}_{t-1|T}^{(j,k)})' \boldsymbol{Z} \\
& + \boldsymbol{Z}' (\boldsymbol{\Sigma}_{t|T}^{k} - \boldsymbol{\Sigma}_{t,t-1|T}^{(j,k)} \hat{\boldsymbol{\mathcal{B}}}_k' - \hat{\boldsymbol{\mathcal{B}}}_k \boldsymbol{\Sigma}_{t,t-1|T}^{(j,k)}{}' + \hat{\boldsymbol{\mathcal{B}}}_k \boldsymbol{\Sigma}_{t-1|T}^{(j,k)} \hat{\boldsymbol{\mathcal{B}}}_k') \boldsymbol{Z} \Big).
\end{aligned}
\tag{53}
$$

Furthermore, for the transition probabilities, we get an equation similar to Equation 10 that looks as follows:

$$
\mathrm{P}[S_t = k | S_{t-1} = j] = \frac{\sum_{t=2}^{T} \mathrm{P}[S_{t-1} = j, S_t = k | \mathcal{I}_T]}{\sum_{t=1}^{T-1} \mathrm{P}[S_t = j | \mathcal{I}_T]}. \tag{54}
$$

We again initialise the algorithm with random numbers for all parameters. We then run the Kim filter and smoother and estimate the new parameters. We keep repeating this until the parameters converge or until we have reached 1000 iterations. Furthermore, to identify the model, we follow Brownlees and Kole (2022) and set the upper part of $\hat{\boldsymbol{\Lambda}}_1$ to an $r \times r$ lower triangular matrix with ones on the diagonal elements, with r again being the number of factors in the model.

To make one-step ahead forecasts, we must first forecast the factors for every possible path of states at time T and $T+1$. Thereafter, we can produce the actual forecast. Hence, we write the following:

$$
\hat{\boldsymbol{\mu}}_{T+1|T}^{(j,k)} = \hat{\boldsymbol{\mathcal{B}}}_k \hat{\boldsymbol{\mu}}_{T|T}^{j}, \tag{55}
$$

$$
\hat{\boldsymbol{y}}_{T+1|T} = \sum_{j=1}^{m} \sum_{k=1}^{m} \mathrm{P}[S_T = j, S_{T+1} = k | \mathcal{I}_T] \big(\hat{\boldsymbol{c}}_k + \hat{\boldsymbol{\Lambda}}_k \hat{\boldsymbol{f}}_{T+1|T}^{(j,k)} \big). \tag{56}
$$

2.4 Model Comparison

The aforementioned models are compared in both a simulation study and an application. The models are compared based on their one-step ahead point forecasts, using the Mean Squared Prediction Error (MSPE) and the corresponding Diebold Mariano (DM) statistic. This statistic reads as follows:

$$
\mathrm{DM} = \frac{\bar{d}}{\hat{\sigma}_d / \sqrt{P}} \sim \mathrm{N}(0,1), \tag{57}
$$

where d is the loss differential, where we use the SPE as loss function. Hence, $d_k = e_{1,k}^2 - e_{2,k}^2$ for $k = T+1, \ldots, T+P$, and $e_i = [e_{i,T+1}, \ldots, e_{i,T+P}]$ is the vector of forecast errors of model i, with P the number of forecasts that are made. Moreover, $\bar{d}$ is the sample mean of d, and $\hat{\sigma}_d$ the sample standard deviation of d.

3 Simulation Study

We first perform a simulation study. Herein, we simulate multiple multivariate time series according to several VAR structures, each having different autoregressive parameters and/or a different number of series. Then, we estimate our models for each of the series and we investigate which model performs best under the different circumstances.

3.1 Simulation Procedure

In our simulation study, we perform eight simulations, each having 100 replications. In every simulation, we generate a certain multivariate time series in every replication. We then estimate the parameters in each of our four models based on those series. Thereafter, we make forecasts, and we calculate the performance of the models based on their MSPEs, which is averaged over all series, and we assess the DM statistics.

In each replication, we generate a multivariate time series of length 1100 based on a VAR model with one lag. Furthermore, we simulate a series representing the state, determined by a Markov process. This series is allowed to take on two values, representing two states, an expansion and a recession. When in state one (that mimics an expansion), the probability to stay in that state is a random number between 0.9 and 1[11]. Moreover, when in state two, the probability to stay in that state is a random number between 0.8 and 0.9. Furthermore, we start in state one. The VAR model has different parameters in the states. We have eight different simulations. In every simulation, we have a different VAR structure. We estimate a VAR with 2, 3, 4, and 5 series, giving us four choices. Moreover, with each of these series, we generate two different VAR models. In the first, we have high positive parameters only in the first state, and intermediate parameters in the second state. In the second model, we have again intermediate parameters, but now in the first state, and low parameters in the second state. More elaboration can be found in Appendix J. Generating data with a VAR with 2, 3, 4, and 5 series enables us to comprehensively compare the models while keeping the research manageable. The different VAR structures are investigated to determine how the models compare to each other under different correlation structures. We expect that the higher the correlation, the better the dynamic factor models perform with respect to the UMSARM, because there is more co-movement between the series. Besides, when we investigate more series there might be more co-movement, and we thus expect that the dynamic factor models perform better relative to the UMSARM as the number of series gets larger.

In each simulation, we rank our models based on their MSPEs and the corresponding DM

[11] The random numbers can be reproduced using Matlab (MathWorks (2021)) with the default random seed.

statistics, which are computed for each replication separately, but with all series combined within a replication. We also list an overall MSPE ratio for each model, which is the average of the MSPEs taken over all replications of the model, divided by the average of the MSPEs taken over all replications of the UMSARM. In each replication, the parameters in the models are estimated based on the first 1000 observations. Thereafter, 100 one-step ahead forecasts are made. Naturally, the dynamic factor models are estimated using all series in a replication, and forecast all series simultaneously. The univariate Markov switching model however, is estimated multiple times within a replication, once for each time series. Also, the time series are forecast separately.

Because this is the first simulation study in this setting, we consider a VAR with only two states and one lag for simplicity. Besides, in our models, we also use two states and one lag. Furthermore, the number of factors that we use in the three dynamic factor models is one for the VARs with two and three series, and two for the VARs with four and five series. Also, we consider some extreme VAR structures, where we do not allow for some positive and some negative numbers within the same autoregressive parameter. Considering more lags, more states, more series, a different number of factors, or a VAR with more nuanced parameters is left for further research.

3.2 Results

In this section, we first discuss the MSPE ratios. Thereafter we draw attention to the results regarding the DM statistics. The MSPE ratios are shown in Table 1.

Table 1: The MSPE ratios for all VAR forms.

Parameter values	Dimension	UMSARM	DFM	DFMM	MSDFM
HIGH	2	1	0.77	0.79	0.74
	3	1	0.79	0.82	0.77
	4	1	0.83	0.94	0.79
	5	1	0.88	0.90	0.80
LOW	2	1	0.90	0.87	0.87
	3	1	0.93	0.94	0.89
	4	1	0.96	0.97	0.90
	5	1	0.94	0.99	0.91

The parameter value HIGH corresponds with the VAR with high parameters in the first state and intermediate parameters in the second state. Similarly, the parameter value LOW corresponds with the VAR with intermediate parameters in the first state, and low parameters in the second state. Besides, the dimension represents the number of series in the VAR.

Generally, when looking at the table, we see that the MSDFM performs the best, followed by the DFM, which is in turn closely followed by the DFMM. The UMSARM performs worst. When the VAR has high parameters in the first state and intermediate parameters in the second state, the dynamic factor models seem to outperform the UMSARM with a larger margin than when the VAR has intermediate parameters in the first state and low parameters in the second state. Moreover, as the number of series in the VAR get larger, the dynamic factor models seem to outperform the UMSARM with a smaller margin.

More specifically, we find for the VAR with two series and high parameters in the first state and intermediate parameters in the second state, that the MSPEs of the DFM and the MSDFM are roughly 25% less than that of the UMSARM, with a standard error of 1.4% for the DFM and 0.7% for the MSDFM, and with the MSDFM being the lowest one with a small but significant margin. Furthermore, we notice that that the MSPE of the DFMM is 20% less than that of the UMSARM, with a standard error of 3.4%. When looking at the VAR with three series and the same distribution of parameters, we see again that the MSDFM performs best, followed by the DFM and then the DFMM, and that UMSARM performs worst. When looking further, we see that when the number of series gets larger, the results do not change drastically. MSDFM has the lowest MSPE, followed by the DFM, then the DFMM, and the UMSARM performs worst. In the VAR with four series, the performance of the DFM is closer to the performance of the MSDFM, which has an MSPE of roughly 20% less than the UMSARM, and in the VAR with five series its performance is closer to that of the DFMM, which has an MSPE of around 10% less than the UMSARM.

When investigating the VARs that have intermediate parameters in the first state and low parameters in the second state, we find that the order of the models with respect to performance does not change but that the overall performance of the three dynamic factor models lie closer to that of the UMSARM. In all VARs, the MSDFM performs best, having an MSPE of roughly 10% less than that of the UMSARM. The MSPE of the DFM ranges between 90% and 96% of that of the UMSARM, and the MSPE of the DFMM ranges between 87% and 99% of that of the UMSARM.

The results regarding the DM statistics are shown in Table 2.

Table 2: The number of DM statistics that are significant on a 5% significance level for all VAR forms.

Parameter values	Dimension	UMSARM - DFM		UMSARM - DFMM		UMSARM - MSDFM		DFM - DFMM		DFM - MSDFM		DFMM - MSDFM	
HIGH	2	1	98	1	98	0	100	6	10	2	22	2	21
	3	1	96	3	94	0	99	9	10	4	20	5	16
	4	7	92	14	77	1	97	26	14	6	26	8	31
	5	9	87	14	72	0	93	28	22	6	29	11	36
LOW	2	3	84	0	83	0	84	3	6	3	9	3	7
	3	7	62	6	60	0	73	13	20	2	18	9	17
	4	11	56	15	55	0	74	19	20	1	21	11	27
	5	7	62	13	49	2	74	32	11	10	20	6	22

The parameter values and dimension mean the same as in Table 1. Moreover, in the columns, models A-B are listed, which means that model A is compared to model B. Also, for each row, a number in the first column below A-B is listed, which is the number of times that model A performs significantly better than model B. Likewise, for each row, a number in the second column below A-B is listed, which is the number of times that model B performs significantly better than model A.

When looking at this table, we find that the MSDFM performs the best, followed by the DFM and the DFMM, which perform equally well. The UMSARM performs worst. The difference in performance between the dynamic factor models and the UMSARM is smaller when the VAR has intermediate parameters in the first state and low parameters in the second state than when the VAR has high parameters in the first state and intermediate parameters in the second state. Moreover, in both parameter distributions, the difference in performance between the dynamic factor models and the UMSARM gets smaller when the number of series gets larger. When comparing the dynamic factor models, both the number of series in the VAR and the parameter distribution does not matter for the difference in performance.

For the VAR with two series and high parameters in the first state and intermediate parameters in the second state, we clearly see that the univariate Markov switching model performs much worse than all three the dynamic factor models. We also notice immediately that the MSDFM performs better than the DFM and the DFMM, however this difference is smaller than the difference between the UMSARM and the dynamic factor models. Regarding the DFM and the DFMM, we observe that the DFMM significantly outperforms the DFM 10 times, the DFM significantly outperforms the DFMM 6 times, and the models perform equally well in roughly 85% of the replications. Looking at the VAR with three series and the same distribution of parameters, we observe similar results. The UMSARM performs again much worse than all three the dynamic factor models, even though the number of DM statistics in favour of the dynamic factor models has become a bit less. We also again see that the MSDFM performs better than both the DFM and the DFMM, which again perform equally well.

When looking further into the table, to the VARs with four or five series, we notice the same patterns. The UMSARM performs much worse than all three the dynamic factor models, but the number of DM statistics in favour of the dynamic factor models becomes less as the number of series gets larger. The MSDFM outperforms both the DFM and the DFMM. Moreover, the number of significant DM statistics when comparing the DFM and the DFMM gets larger, but the difference in the number of DM statistics in favour of the former and the number of DM statistics in favour of the latter stays relatively small.

When the VARs have intermediate parameters in the first state and low parameters in the second state, we find that the dynamic factor models still outperform the UMSARM, but that the number of significant DM statistics is a lot less than it was before. The number of significant DM statistics in favour of the MSDFM when comparing it to the DFM and the DFMM does not change that much, and we again find that the MSDFM performs best. Again, for each number of series, the difference in the number of DM statistics in favour of the DFM and the number of DM statistics in favour of the DFMM stays relatively small. As in the VAR with high parameters in the first state and intermediate parameters in the second state, we find that when comparing the dynamic factor models with the UMSARM the number of DM statistics in favour of the dynamic factor models becomes less as the number of series gets larger.

Summarising the results of both the MSPE ratios and the DM statistics, we find the following. The MSDFM clearly performs best, and the UMSARM performs worst, for both parameter distributions and for all number of series. Furthermore, there is only a tiny difference between the DFM and the DFMM, which is in favour of the DFM. When the parameters in the VAR are large, we find that the dynamic factor models perform much better with respect to the UMSARM than when the VAR parameters are small, which is just as expected. Remarkable is that when the number of parameters gets larger, the difference in performance between the UMSARM and the dynamic factor models becomes smaller, for both parameter distributions, which is different than expected. We thus conclude that if data originates from a VAR model, the MSDFM has the best forecasting performance across the models used in this research. Hence, if one examines a data set and believes that the data originates from a VAR, but there are not enough observations to estimate a VAR, the MSDFM should be looked into.

4 Empirical Application

As an application, we compare the forecast performance of the aforementioned models, using a data set including three macroeconomic variables, annual GDP growth, unemployment rate, and inflation rate, for a total of seven countries, Canada, France, Germany, Italy, Japan, the

UK, and the USA, better known as the group of seven (or G7).

When using a VAR for all countries for only one variable and one state only, we would already need 56 parameters. If we wish to use more variables or allow for more states, the number of parameters really explodes. Hence, to properly analyse this data set, we must make use of the aforementioned models. We analyse different sets of series within this data set and for each we investigate which of our considered models works best.

4.1 Data

As mentioned before, we use a data set including three macroeconomic variables for a total of seven countries. The data set includes quarterly data in the period from 1991Q1 until 2021Q4, and is gathered from the Organisation for Economic Cooperation and Development (OECD). Some descriptive statistics of the variables are shown in Table 3. The table consists of three parts, where the first part includes all the observations, the second part includes observations in recessions only, as indicated by the National Bureau of Economic Research (NBER), and the third part includes observations in expansions only. To distinguish between recessions and expansions, we regarded a whole quarter as being in a recession (expansion) when most of its months were in a recession (expansion). Worth to mention is that the NBER indicated recessions are recessions occurring in the USA. However, since the economy of the USA is the largest in the world and other economies co-move with this economy, we believe using solely these recessions is worthwhile for data exploration.

Table 3: Descriptive statistics of the macroeconomic variables for all countries from 1991Q1 until 2021Q4 **(left)**, in recessions only **(middle)**, and in expansions only **(right)**.

Variable	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
GDP growth Canada	0.56	1.46	-0.55	1.11	0.67	1.45
GDP growth France	0.39	2.27	-0.79	1.76	0.51	2.29
GDP growth Germany	0.33	1.51	-0.51	1.83	0.41	1.46
GDP growth Italy	0.18	2.06	-1.04	2.12	0.30	2.03
GDP growth Japan	0.19	1.28	-0.69	1.80	0.28	1.20
GDP growth UK	0.50	2.49	-0.65	1.12	0.61	2.56
GDP growth USA	0.62	1.22	-0.47	0.84	0.72	1.20
Unemployment rate Canada	7.86	1.61	7.25	1.27	7.92	1.63
Unemployment rate France	9.78	1.46	8.27	0.76	9.92	1.43
Unemployment rate Germany	6.90	2.38	6.92	1.42	6.89	2.46
Unemployment rate Italy	9.71	1.73	7.90	1.04	9.89	1.68
Unemployment rate Japan	3.75	0.99	4.14	1.07	3.71	0.98
Unemployment rate UK	6.39	1.83	5.89	1.22	6.44	1.88
Unemployment rate USA	5.91	1.74	5.99	1.66	5.90	1.75
Inflation rate Canada	1.93	1.13	2.40	1.69	1.89	1.06
Inflation rate France	1.49	0.83	1.95	1.18	1.44	0.78
Inflation rate Germany	1.80	1.22	2.03	0.93	1.78	1.25
Inflation rate Italy	2.30	1.64	2.78	1.63	2.25	1.64
Inflation rate Japan	0.35	1.11	0.56	1.50	0.33	1.07
Inflation rate UK	2.31	1.34	3.09	2.01	2.24	1.25
Inflation rate USA	2.39	1.23	2.68	2.08	2.36	1.12

Regarding GDP growth, we observe that on average, the USA had the largest growth (0.62%), followed closely by Canada (0.56%). The lowest average growth during this period was in Italy (0.18%) and Japan (0.19%). The highest volatility is attained in the UK (2.49). Noteworthy is that the volatility over the whole period for the UK is more than twice as high as during recessions. When looking at the unemployment rates, we notice that Japan performs really well, having an average unemployment rate of only 3.75%. The worst performing country is France, with an average unemployment rate of 9.78%. Japan is also the country with the lowest average inflation rate (0.35%), which is even more than four times as small as the average inflation rate of France, the country with the second lowest average inflation rate (1.49%). The highest average inflation rate was in the USA (2.39%).

To give more insights in the data, we show a plot of the GDP growth, inflation rate, and unemployment rate in the USA in Figure 1.

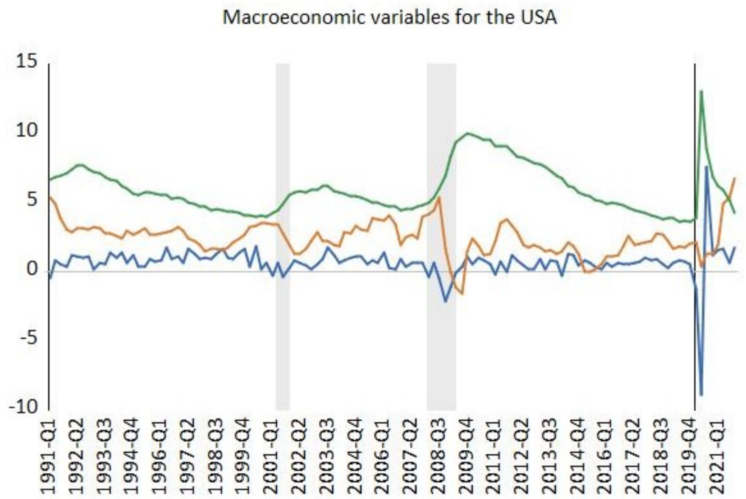

Figure 1: Plot of the macroeconomic variables for the USA from 1991Q1 until 2021Q4. The blue line represents GDP growth, the orange line the inflation rate, and the green line the unemployment rate. The two grey areas are NBER indicated recession periods that lasted longer than one quarter. The two black vertical lines are NBER indicated recessions that lasted only one quarter.

We see that all variables do not show a trend over time. We do however clearly see that the unemployment rate rises quickly in recessions and slowly declines in expansions. For the inflation rate, we see that it declines in recessions and rises in expansions, but both roughly at the same speed. GDP growth stays quite the same during expansions, but in recessions it declines quickly in the first half and rises quickly in the second half, except for the last recession, where it just declines quickly, and rises in the expansion right after.

Furthermore, we list the correlations between the GDP growth in all countries in Table 4.

Table 4: The correlations between the GDP growth in all countries during the whole period.

	Canada	France	Germany	Italy	Japan	UK
France	0.88					
Germany	0.79	0.84				
Italy	0.84	0.97	0.87			
Japan	0.67	0.62	0.71	0.65		
UK	0.89	0.93	0.86	0.92	0.71	
USA	0.90	0.86	0.77	0.84	0.68	0.91

All correlations are significant on a 1% significance level.

All correlations are positive, meaning that the GDP growth in all countries rise together and decline together. The weakest correlations are between Japan and the other countries. The strongest is that between Italy and France, which is 0.97.

We also list the correlations between the GDP growth in all countries in recessions and in expansion only, this is shown in Table 5.

Table 5: The correlations between the GDP growth in all countries during recessions only **(lower)** and during expansions only **(upper)**.

	Canada	France	Germany	Italy	Japan	UK	USA
Canada		0.89	0.82	0.86	0.71	0.90	0.90
France	0.64**		0.87	0.97	0.68	0.93	0.88
Germany	0.41	0.52		0.88	0.68	0.89	0.80
Italy	0.50	0.92***	0.74***		0.68	0.93	0.86
Japan	0.18	0.11	0.78***	0.33		0.75	0.70
UK	0.62**	0.80***	0.63**	0.83***	0.39		0.92
USA	0.66**	0.54*	0.47	0.55*	0.40	0.75***	

The lower triangular matrix represents the correlations during recessions and the upper triangular matrix depicts the correlations during expansions. For the lower triangular matrix, significance is represented by one, two, or three asterisks, meaning whether the correlation is significant at the 10%, 5%, or 1% significance level, respectively. For the upper triangular matrix, the significance is not explicitly indicated in the table because all correlations are significant on a 1% significance level.

During expansions, we observe the same as in the whole sample. All correlations are positive, and the weakest correlations are those between Japan and all other countries. The strongest is again that between Italy and France, which is again 0.97. During recessions, not every correlation is significant anymore, of which most are correlations with Japan. On average, all correlations are lower than during expansions, with the highest correlation now being 0.92 between Italy and France, and the lowest being 0.11 between Japan and France, which was 0.68 during expansions and 0.62 during the whole period.

Additionally, we show the correlations between the GDP growth and the lag of the GDP growth in all countries, this is done in Table 6.

Table 6: The correlations between the GDP growth and the GDP growth with one lag in all countries during the whole period.

	Canada	France	Germany	Italy	Japan	UK	USA
Lag Canada	-0.12	-0.45***	-0.22**	-0.40***	-0.09	-0.27***	-0.17*
Lag France	0.00	-0.31***	-0.07	-0.27***	0.06	-0.12	-0.05
Lag Germany	-0.09	-0.38***	-0.20**	-0.35***	-0.08	-0.27***	-0.15
Lag Italy	0.05	-0.26***	-0.02	-0.22**	0.09	-0.09	-0.01
Lag Japan	-0.14	-0.38***	-0.18**	-0.32***	-0.14	-0.30***	-0.16*
Lag UK	-0.17**	-0.48***	-0.24***	-0.44***	-0.11	-0.33***	-0.22**
Lag USA	-0.12	-0.43***	-0.17**	-0.37***	-0.02	-0.25***	-0.16**

The addition of one, two, or three asterisks means significance on the 10%, 5%, and 1% significance level respectively.

We notice that all significant correlations are negative, with the correlation between the GDP growth in France and the lag of the GDP growth in the UK being the one the one with the largest magnitude (-0.48). Meaning that if the GDP growth in the UK at some period in time declines (rises), the GDP growth in France generally rises (declines) the period after that. Furthermore, out of the seven countries, five are significantly correlated with their own lag, and we see quite a lot significant correlations between the GDP growth in the countries and the lags of the GDP growth in other countries, except for Canada and Japan.

Ultimately, we show the scree plot of the Principal Component Analysis (PCA) of all variables and all countries combined in Figure 2

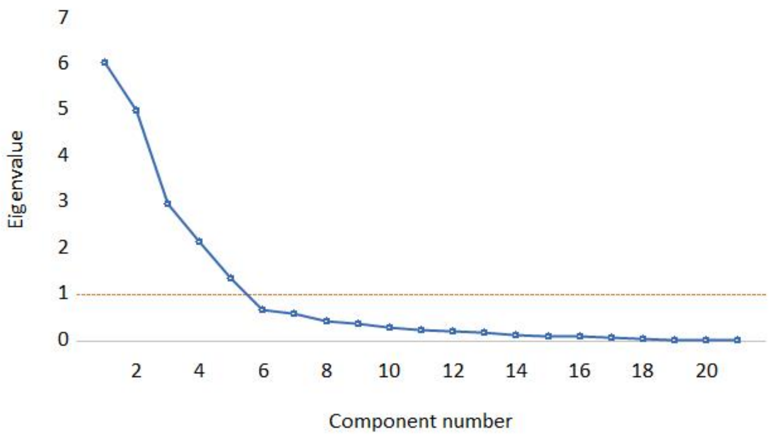

Figure 2: Scree plot of the PCA of all variables and all countries combined.

We find that the sixth component corresponds with both the first eigenvalue that is lower than one and the so-called elbow in the scree plot. When retaining five components, we find that 84% of the variance is explained by the principal components. The scree plots of the other subsets of the data can be found in Appendix K.

4.2 Methodology

In our application, we consider a total of eleven subsets of our data set. Then, similar as in the simulation study, we estimate, on each subset, each of our models, and we measure their performance by their MSPEs and the corresponding DM statistics, averaged over all considered series. The eleven subsets are the following. For each country separately, we consider all variables, giving us the first seven subsets (of three series each). Then, for each variable separately, we consider all countries, giving us three more (of seven series each). The last is the full data set, including all countries and all variables (which has 21 series). For the subsets with three series, we use one factor in the dynamic factor models. Although the PCA suggest retaining two or three factors, we decide to use only one because the dynamic factor models are used to reduce the number of parameters in the model, and this is not achieved otherwise[12]. For the subsets with seven series, we use three factors, as suggested by the PCA. For the unemployment rate, we find that the first three eigenvalues are larger than one, and when retaining three components, 84% of the variance is explained. For the inflation rate, we find that only the first eigenvalue is larger

[12] When using a VAR with three series, we have 21 parameters. When using a dynamic factor model with three series and two factors, we have 26 parameters, and with three factors even more. When using one factor, we need only 17 parameters.

than one, but it only explains 62% of the variance. When retaining three components, 87% of the variance is explained. For the GDP growth rate, again only the first eigenvalue is larger than one, and for this one explains 84% of the variance. However, a clear interpretation can be associated with the second and third component, with the second component being solely the GDP growth of Japan, and the third component being the difference in GDP growth between North America (Canada and the USA) and Europe (France, Germany, and Italy). For the full data set, we use and five factors, again as suggested by the PCA, which suggests retaining five components as explained in Section 4.1. We expect, at least for the subset with GDP, that using a dynamic factor model yields additional performance relative to the UMSARM, because a lot of countries have GDP growth that is significantly correlated with the lag of the GDP growth in other countries. Moreover, based on the simulation study, we expect that the addition of dynamic factors provides less improvement as the number of series grows.

As in the simulation study, we rank our models based on their MSPEs and the corresponding DM statistics, which are combined for all series. Here, we again list an overall MSPE ratio for each model. Again, the dynamic factor models are estimated using all investigated series, and forecast all series simultaneously, and the univariate Markov switching model is estimated once for each time series and the series are forecast separately.

In the application, the estimation sample is an expanding window. First, the models are estimated based on the data from 1991Q1 until 2011Q4. Thereafter, one-step ahead forecasts are made until 2013Q4. Then, the models are estimated based on the data from 1991Q1 until 2013Q4, and one-step ahead forecasts are made until 2015Q4, and so forth until we obtain a forecast for 2021Q4. Since the number of parameters in the models can get large, the expanding window is chosen to obtain a reasonable ratio of number of observations and number of parameters that need to be identified. An issue with the expanding window however, is that in combination with nested models, the DM test might not be valid. Hence, the DM statistics obtained when comparing the DFM with either the DFMM or the MSDFM must be interpreted with caution.

Since this study is among the first to examine the Markov switching dynamic factor model and the dynamic factor mixture model in an applied setting, we use two states and one lag in our models for simplicity. Again, considering more lags, more states, or a different number of factors is left for further research. Also, further research might consider using a rolling window rather than an expanding window.

4.3 Results

In this section, we discuss the MSPE ratios and the DM statistics of all subsets of our data set. These are shown in Table 7.

Table 7: The MSPE ratios and DM statistics for all subsets of the data.

	MSPEs				DM statistics					
	UMSARM	DFM	DFMM	MSDFM	UMSARM - DFM	UMSARM - DFMM	UMSARM - MSDFM	DFM - DFMM	DFM - MSDFM	DFMM - MSDFM
GDP	1	1.33	0.59	1.06	-2.43**	2.32**	-0.68	2.68***	2.12**	-2.55**
Inflation	1	3.67	NA	2.77	-5.18***	NA	-3.87***	NA	1.58	NA
Unemployment	1	3.12	NA	13.36	-4.37***	NA	-4.66***	NA	-3.92***	NA
Canada	1	0.75	0.81	0.81	-0.68	0.68	0.46	-0.61	-0.44	0.03
France	1	0.97	0.63	0.68	0.09	0.98	0.88	2.63***	2.37**	-2.64***
Germany	1	1.16	2.75	1.14	-0.49	-1.19	-1.80*	-1.03	0.07	1.10
Italy	1	1.45	2.14	1.45	-2.18**	-3.14***	-1.46	-1.51	-0.02	1.30
Japan	1	1.04	0.71	0.77	-0.17	0.73	0.56	1.77*	1.31	-0.35
UK	1	0.72	0.64	0.76	0.89	1.15	0.75	2.83***	-0.68	-1.68*
USA	1	1.14	0.89	1.11	-0.55	0.37	-0.51	2.20**	0.38	-1.72*
All	1	0.82	NA	NA	0.97	NA	NA	NA	NA	NA

For the DM statistics, A-B means we are comparing models A and B, where a negative number means that model A performs better, and a positive number means that model B performs better. In the rows we have the subsets of the data, where for instance GDP means the subset with GDP growth for all countries, and Canada means the subset of all series for Canada only. Besides, in the last row we have all series for all countries. For some subsets of the data, the parameters could not be identified for all models, these MSPE ratios and corresponding DM statistics are listed as NA. Furthermore, the addition of one, two, or three asterisks means significance on the 10%, 5%, and 1% significance level respectively.

The first thing we notice when looking at this table is that the DFMM and the MSDFM cannot be utilised for every subset of the data. This could either be due to certain properties of the data or due to the ratio of number of observations and number of parameters that need to be identified. When looking further into the table, we see that if we are able to estimate all models, the UMSARM and the DFMM perform best, followed by the MSDFM, and the DFM performs worst. If we study the table in more detail, we find that the UMSARM performs best when looking at one variable only but for each country simultaneously, which is different than expected. The parameters of the DFMM cannot be identified in two out of three times, and the DFM and the MSDFM perform much worse than the UMSARM, with MSPE ratios of 1.33, 3.67, and 3.12 for the DFM, and 1.06, 2.77, and 13.36 for the MSDFM, for GDP growth, inflation, and unemployment respectively, of which five have highly significant DM statistics. When we look at all three variables for one country, we find that most DM statistics are insignificant. When comparing the UMSARM to the three dynamic factor models, we find only three significant DM statistics, where the UMSARM performs significantly better than the MSDFM when forecasting the series in Germany, and it performs significantly better than both the DFM and the DFMM when forecasting the series in Italy. If we compare the dynamic factor models with each other, the DFM gets significantly outperformed by the DFMM four times, and once by the MSDFM.

However, as mentioned in Section 4.2, we cannot draw clear conclusions from this. Furthermore, the MSDFM is significantly outperformed by the DFMM three times. Lastly, if we look at all countries and all variables at once, the parameters in the DFMM and the MSDFM cannot be identified, and the MSPE ratio is 0.82 for the DFM, but this is not a significant difference. The decrease of the performance of the dynamic factor models with respect to the UMSARM as the number of series gets larger is comparable to that of the simulation study, since the relative performance is better when there are only three series than when there are seven series. However, this relation does not hold when the number of series is 21, but this might mean that the last mentioned subset can be compared to the higher parameters in the simulation study and the other subsets to the lower parameters in the simulation study.

In short, the parameters in the DFMM and the MSDFM cannot always be identified, but when they are, the DFMM performs adequately. The DFM and MSDFM perform relatively well in general, but occasionally performs much worse than the UMSARM, which is, for this data, the safest option across the used models. Hence, when examining macroeconomic variables, we conclude that using the UMSARM is the best option. Worth to mention is that adjusting the number of factors in the dynamic factor models might increase the model performance and might make it possible to identify all parameters in each model.

4.4 Robustness

To check whether our findings of the application are robust, we investigate the same eleven subsets of our data again, but with the observations during the COVID-19 pandemic omitted. Hence, we obtain forecasts until 2019Q4 instead of 2021Q4. Then, we again rank our models based on their MSPEs and the corresponding DM statistics, which are combined for all series. Here, we again list an overall MSPE ratio for each model. The results can be found in Table 8.

Table 8: The MSPE ratios and DM statistics for all subsets of the data with the observations during the COVID-19 pandemic omitted.

	MSPEs				DM statistics					
	UMSARM	DFM	DFMM	MSDFM	UMSARM - DFM	UMSARM - DFMM	UMSARM - MSDFM	DFM - DFMM	DFM - MSDFM	DFMM - MSDFM
GDP	1	1.01	1.29	1.11	-0.19	-1.59	-1.80*	-1.46	-1.78*	0.92
Inflation	1	5.51	NA	3.73	-4.39***	NA	-3.01***	NA	1.38	NA
Unemployment	1	37.54	NA	166.29	-8.30***	NA	-4.01***	NA	-3.09***	NA
Canada	1	1.85	1.78	1.57	-2.80***	-2.86***	-2.87***	0.49	1.43	1.58
France	1	36.35	4.13	6.58	-2.81***	-4.29***	-6.10***	2.54**	2.39**	-3.56***
Germany	1	14.21	4.94	4.71	-9.02***	-5.43***	-5.50***	7.03***	6.29***	0.26
Italy	1	23.98	20.81	21.27	-8.32***	-6.74***	-3.95***	2.20**	0.48	-0.09
Japan	1	2.29	1.84	1.78	-3.24***	-2.92***	-2.81***	1.90*	2.15**	1.58
UK	1	11.05	4.63	5.39	-5.90***	-5.95***	-5.07***	3.70***	3.66***	-0.86
USA	1	9.22	5.00	7.00	-4.55***	-3.87***	-4.21***	2.67***	1.32	-2.77***
All	1	5.42	NA	NA	-11.83***	NA	NA	NA	NA	NA

In this table, the rows and columns have the same meaning as in Table 7. Also, again the addition of one, two, or three asterisks means significance on the 10%, 5%, and 1% significance level respectively.

Noticeable in this table is that the MSPE ratios of the dynamic factor models are much higher than in Table 7, and it is clear to see that the UMSARM outperforms all other models, since all MSPE ratios for the dynamic factor models are large and all the DM statistics in the columns with the UMSARM, except for the first row, are in favour of the UMSARM and significant on the 1% significance level. Hence, the co-movement between the series was much larger during the COVID-19 pandemic than before, such that modelling the series simultaneously becomes abundant when looking at the data before this period. When comparing the three dynamic factor models, we find that the DFM generally performs best when looking at all countries but one variable only, and that both the DFMM and the MSDFM perform best when looking at all variables but one country only, although this might be the results of non-valid DM statistics. Since the parameters in both the DFMM and the MSDFM cannot be identified when looking at all countries and all variables, the DFM is the best performing model in this case.

Summarising the results, we find that the models do not behave the same when excluding 2020 and 2021 as when we include those years. The dynamic factor models perform much worse compared to the UMSARM than it did before. In addition, when comparing the dynamic factor models with each other, the results are not quite the same as before. Howbeit, the UMSARM performs the best, which seems to be a robust finding.

5 Conclusion

This paper assesses and compares the performance of the univariate Markov switching autoregressive Model (UMSARM), the dynamic factor model (DFM), the dynamic factor mixture model (DFMM), and the Markov switching dynamic factor model (MSDFM). This is done in both a simulation study and in an applied setting.

In the simulation study, we compare one-step ahead point forecasts of our models on multiple generated data sets according to different vector autoregressive (VAR) structures. All VARs have one lag, but all have different autoregressive parameters and/or a different number of series. Moreover, there are two states in which the VAR has different parameters, and there is a Markov process determining the switching between the states.

We find that the MSDFM performs the best by quite a large proportion. The second best model is the DFM, closely followed by the DFMM. Ultimately, the UMSARM performs worst. Nonetheless, when the parameters in the VAR become smaller values, we see that the difference in performance between the three dynamic factor models and the UMSARM shrinks as well. Besides, the difference in performance between the dynamic factor models and the UMSARM gets smaller as the number of series in the VAR gets larger. When comparing the dynamic factor models to each other, both the number of series in the VAR and the size of the parameters does not matter for the difference in performance.

As an application, we consider a data set including annual GDP growth, unemployment rate, and inflation rate, all for Canada, France, Germany, Italy, Japan, the UK, and the USA, in the period from 1991Q1 until 2021Q4. To compare the one-step ahead point forecast performance of our models, we use eleven subsets of the data set. Finally, we perform a robustness check doing the analysis with the same data subsets, but with the observations during the COVID-19 pandemic omitted (2020Q1-2021Q4). In both the simulation study and the application, the forecasts are compared according to their MSPEs and the corresponding DM statistics.

For the application we find that the UMSARM performs the best, and that the DFM performs worst. The DFMM and the MSDFM perform relatively well, but their parameters cannot be identified at all times. The last mentioned phenomenon mostly happens when the number of series gets large. When the observations during the COVID-19 pandemic are omitted, we again find that the UMSARM performs the best. Moreover, the dynamic factor models perform much worse compared to the UMSARM than they did before.

From these results, we conclude the following. Across the models used throughout this paper, the MSDFM has the best forecasting performance when the data originates from a VAR model. Hence, if one examines a data set and believes that the data originates from a VAR, but there are not enough observations to estimate a VAR, the MSDFM might be worthwhile to look into. Moreover, the DFM and DFMM also perform relatively well when the data originates from a VAR, and the UMSARM performs worst. When the number of series that are investigated gets larger, the performance of the UMSARM improves with respect to the dynamic factor models. When examining macroeconomic variables, we conclude that using the UMSARM is the best option, since the addition of dynamic factors does not yield any additional performance.

However, in this paper, a specific number of factors is used for the dynamic factor models. Adjusting the number of factors might make it possible to identify all parameters in each model. Also, this study assumes the state at time $t + 1$ equals the state at time t in the DFMM. It might be worthwhile to investigate whether a more sophisticated model for the state increases the model performance of the DFMM, as this model misses the transitions. Additionally, to broaden the scope of the research, the simulation study could be augmented with VARs that have more nuanced parameters or VARs that have a larger number of series. To expand the application, other types of data sets could be used, and data sets with more observations could be examined to be able to estimate all the models properly. Lastly, considering a rolling window rather than an expanding window to estimate the parameters might give more insights in the comparison of the DFM with the DFMM and the MSDFM.

References

Akay, O. O., Senyuz, Z., and Yoldas, E. (2013). Hedge fund contagion and risk-adjusted returns: A Markov-switching dynamic factor approach. *Journal of Empirical Finance*, 22:16–29.

Alloza, M. (2017). Is fiscal policy more effective in uncertain times or during recessions? Working paper, Banco de España.

Bai, J. and Wang, P. (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics*, 33(2):221–240.

Brownlees, C. and Kole, E. (2022). High-dimensional dynamic factor models with Markov switching. Working paper, Erasmus University Rotterdam.

Carstensen, K., Heinrich, M., Reif, M., and Wolters, M. H. (2020). Predicting ordinary and severe recessions with a three-state Markov-switching dynamic factor model: An application to the German business cycle. *International Journal of Forecasting*, 36(3):829–850.

Chauvet, M. and Piger, J. (2008). A comparison of the real-time performance of business cycle dating methods. *Journal of Business & Economic Statistics*, 26(1):42–49.

Chauvet, M. and Potter, S. (2013). Forecasting output. *Handbook of Economic Forecasting*, 2:141–194.

Gonçalves, S., Herrera, A. M., Kilian, L., and Pesavento, E. (2022). When do state-dependent local projections work? Working paper, Federal Reserve Bank of Dallas.

Goodwin, T. H. (1993). Business-cycle analysis with a Markov-switching model. *Journal of Business & Economic Statistics*, 11(3):331–339.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2):357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45(1-2):39–70.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

Kim, C.-J. and Nelson, C. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics*, 80(2):188–201.

Kim, M.-J. and Yoo, J.-S. (1995). New index of coincident indicators: A multivariate Markov switching factor model approach. *Journal of Monetary Economics*, 36(3):607–630.

Majumdar, R. and Majumdar, S. (2019). On the conditional distribution of a multivariate normal given a transformation–the linear case. *Heliyon*, 5(2):e01136.

MathWorks (2021). Version 2021a. URL `https://www.mathworks.com/products.html?s_tid=gn_ps`.

Molenaar, P. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, 50(2):181–202.

Nakajima, J. et al. (2011). Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications. Working paper, Institute for Monetary and Economic Studies, Bank of Japan.

Ng, V., Engle, R. F., and Rothschild, M. (1992). A multi-dynamic-factor model for stock returns. *Journal of Econometrics*, 52(1-2):245–266.

Sims, C. A., Waggoner, D. F., and Zha, T. (2008). Methods for inference in large multiple-equation Markov-switching models. *Journal of Econometrics*, 146(2):255–274.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbook on Economic Forecasting. Oxford University Press*, 2:35–59.

Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. *Handbook of Macroeconomics*, 2:415–525.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of Econometrics*, 4:2843–2915.

Appendix

A Derivation of the update step of the Hamilton filter (Hamilton (1989))

$$
\begin{aligned}
\hat{\boldsymbol{\xi}}_{t|t} &= \begin{bmatrix} \mathrm{P}[S_t = 1|\mathcal{I}_t] \\ \vdots \\ \mathrm{P}[S_t = m|\mathcal{I}_t] \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{P}[S_t = 1|\mathcal{I}_{t-1}, y_t] \\ \vdots \\ \mathrm{P}[S_t = m|\mathcal{I}_{t-1}, y_t] \end{bmatrix} \\
&= \frac{1}{f(y_t|\mathcal{I}_{t-1})} \begin{bmatrix} f(S_t = 1, y_t|\mathcal{I}_{t-1}) \\ \vdots \\ f(S_t = m, y_t|\mathcal{I}_{t-1}) \end{bmatrix} \\
&= \frac{1}{f(y_t|\mathcal{I}_{t-1})} \begin{bmatrix} f(y_t|S_t = 1, \mathcal{I}_{t-1})\mathrm{P}[S_t = 1|\mathcal{I}_{t-1}] \\ \vdots \\ f(y_t|S_t = m, \mathcal{I}_{t-1})\mathrm{P}[S_t = m|\mathcal{I}_{t-1}] \end{bmatrix} \\
&= \frac{1}{f(y_t|\mathcal{I}_{t-1})} \begin{bmatrix} f(y_t|S_t = 1) \\ \vdots \\ f(y_t|S_t = m) \end{bmatrix} \odot \hat{\boldsymbol{\xi}}_{t|t-1} \\
&= \frac{\begin{bmatrix} f(y_t|S_t = 1) \\ \vdots \\ f(y_t|S_t = m) \end{bmatrix} \odot \hat{\boldsymbol{\xi}}_{t|t-1}}{\begin{bmatrix} 1 \cdots 1 \end{bmatrix} \left(\begin{bmatrix} f(y_t|S_t = 1) \\ \vdots \\ f(y_t|S_t = m) \end{bmatrix} \odot \hat{\boldsymbol{\xi}}_{t|t-1} \right)}
\end{aligned}
$$

B Derivation of the inner expectation of the Kim smoother (Hamilton (1989))

Without loss of generality, we can write that ξ_{t+1} has a one of position k and zeroes on all other positions, and that we have only two states. Now we can write the following:

$$
\begin{aligned}
\mathbb{E}[\xi_t|\xi_{t+1}, \mathcal{I}_T] &= \begin{bmatrix} \mathrm{P}[S_t = 1|S_{t+1} = k, \mathcal{I}_T] \\ \mathrm{P}[S_t = 2|S_{t+1} = k, \mathcal{I}_T] \end{bmatrix} \\
&\approx \begin{bmatrix} \mathrm{P}[S_t = 1|S_{t+1} = k, \mathcal{I}_t] \\ \mathrm{P}[S_t = 2|S_{t+1} = k, \mathcal{I}_t] \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{P}[S_t = 1, S_{t+1} = k|\mathcal{I}_t] \\ \mathrm{P}[S_t = 2, S_{t+1} = k|\mathcal{I}_t] \end{bmatrix} /\mathrm{P}[S_{t+1} = k|\mathcal{I}_t] \\
&= \begin{bmatrix} \mathrm{P}[S_t = 1|\mathcal{I}_t] \\ \mathrm{P}[S_t = 2|\mathcal{I}_t] \end{bmatrix} \odot \begin{bmatrix} \mathrm{P}[S_{t+1} = k|S_t = 1] \\ \mathrm{P}[S_{t+1} = k|S_t = 2] \end{bmatrix} /\mathrm{P}[S_{t+1} = k|\mathcal{I}_t] \\
&= \hat{\boldsymbol{\xi}}_{t|t} \odot \left[\begin{bmatrix} p_{k1} \\ p_{k2} \end{bmatrix} /(\hat{\boldsymbol{\xi}}_{t+1|t})_k \right] \\
&= \hat{\boldsymbol{\xi}}_{t|t} \odot [P' \boldsymbol{e}_k/(\hat{\boldsymbol{\xi}}_{t+1|t})_k], \text{ with } \boldsymbol{e}_k \text{ a unit vector with a one at position } k \\
&\qquad\qquad \text{and zeroes on all other positions.} \\
&= \hat{\boldsymbol{\xi}}_{t|t} \odot P'[\boldsymbol{e}_k \oslash \hat{\boldsymbol{\xi}}_{t+1|t}] \\
&= \hat{\boldsymbol{\xi}}_{t|t} \odot \boldsymbol{P}'[\boldsymbol{\xi}_{t+1} \oslash \hat{\boldsymbol{\xi}}_{t+1|t}].
\end{aligned}
$$

C Derivation of p^*_{ijt} in the E-step of the EM algorithm of the Markov switching autoregressive model (Hamilton (1990))

$$
\begin{aligned}
p^*_{ijt} &= \mathrm{P}[S_t = i, S_{t-1} = j | \mathcal{I}_T] \\
&= \mathrm{P}[S_{t-1} = j | S_t = i, \mathcal{I}_T]\mathrm{P}[S_t = i | \mathcal{I}_T] \\
&= \mathrm{P}[S_{t-1} = j | S_t = i, \mathcal{I}_{t-1}]\mathrm{P}[S_t = i | \mathcal{I}_T] \\
&= \frac{\mathrm{P}[S_t = i, S_{t-1} = j | \mathcal{I}_{t-1}]\mathrm{P}[S_t = i | \mathcal{I}_T]}{\mathrm{P}[S_t = i | \mathcal{I}_{t-1}]} \\
&= \frac{p_{ij}\mathrm{P}[S_{t-1} = j | \mathcal{I}_{t-1}]\mathrm{P}[S_t = i | \mathcal{I}_T]}{\mathrm{P}[S_t = i | \mathcal{I}_{t-1}]} \\
&= \frac{(\boldsymbol{P})_{ij} \times (\hat{\boldsymbol{\xi}}_{t|T})_i \times (\hat{\boldsymbol{\xi}}_{t-1|t-1})_j}{(\hat{\boldsymbol{\xi}}_{t|t-1})_i} \\
&= \frac{\boldsymbol{P} \odot (\hat{\boldsymbol{\xi}}_{t|T}\hat{\boldsymbol{\xi}}'_{t-1|t-1})_{ij}}{(\hat{\boldsymbol{\xi}}_{t|t-1})_i} \\
&= \left[\boldsymbol{P} \odot (\hat{\boldsymbol{\xi}}_{t|T}\hat{\boldsymbol{\xi}}'_{t-1|t-1}) \oslash (\hat{\boldsymbol{\xi}}_{t|t-1}[1 \cdots 1])\right]_{ij}
\end{aligned}
$$

D Derivation of parameter expressions in the M-step of the EM algorithm of the Markov switching autoregressive model (Hamilton (1990))

When optimising 7 over the initial distribution, we need to solve

$$\frac{\mathrm{d}}{\mathrm{d}\rho_k}\sum_{j=1}^{n}\left[p^*_{j1}\log\bigl(\rho_j\bigr)-\kappa\left(\sum_{j=1}^{n}\rho_j-1\right)\right]=0,$$

with κ a Lagrange multiplier, enforcing $\sum_j \rho_j = 1$. This leads to $\frac{p^*_{k1}}{\rho_k} - \kappa = 0$. Hence, $p^*_{k1} = \kappa\rho_k$. $\sum_k p^*_{k1} = \sum_k \rho_k$, implying that κ equals one, and thus $p^*_{k1} = \rho_k$. Hence, by definition of p^*_{k1}, $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{\xi}}_{1|T}$.

When optimising 7 with respect to the transition probabilities, we need to solve

$$\frac{\partial}{\partial p_{kl}}\left[\sum_{t=2}^{T}\left[\sum_{i,j=1}^{n}p^*_{ijt}\log\bigl(f_i(y_t|\boldsymbol{\theta_i})p_{ij}\bigr)\right]-\kappa\left(\sum_{i=1}^{n}p_{il}-1\right)\right]=0,$$

with κ again a Lagrange multiplier, now enforcing $\sum_i {}_{il} = 1$. This leads to $\sum_{t=2}^{T}\frac{p^*_{klt}}{p_{kl}} - \kappa = 0$. Hence, $\sum_{t=2}^{T} p^*_{klt} = \kappa p_{kl}$. Since $\sum_k p_{kl} = 1$, $\sum_{t=2}^{T}\sum_k p^*_{klt} = \kappa$. Because $\sum_k p^*_{klt} = p^*_{l(t-1)}$, $\kappa = \sum_{t=2}^{T} p^*_{l(t-1)}$, and $\hat{p}_{kl} = \frac{\sum_{t=2}^{T} p^*_{klt}}{\sum_{t=2}^{T} p^*_{l(t-1)}}$.

When optimising 7 with respect to the shape parameters $\boldsymbol{\theta}_k$, we need to solve

$$\frac{\partial}{\partial\theta_k}\left[\sum_{t=2}^{T}\left[\sum_{i,j=1}^{n}p^*_{ijt}\log\bigl(f_i(y_t|\boldsymbol{\theta_i})p_{ij}\bigr)\right]\right]=\sum_{t=2}^{T}\left[\sum_{j=1}^{n}p^*_{kjt}\frac{\partial\log\bigl(f_k(y_t|\boldsymbol{\theta_k})\bigr)}{\partial\theta_k}\right]=0.$$

Hence, we have a weighted version of our standard maximum likelihood results, and we can write the following:

$$\hat{\boldsymbol{\beta}}_k = (\boldsymbol{X}_l' P_k^* \boldsymbol{X}_l)^{-1}\boldsymbol{X}_l' P_k^* \boldsymbol{y},$$
$$\hat{\sigma}_k^2 = \frac{\sum_{t=2}^{T} p^*_{kt}(y_t - \hat{\boldsymbol{\beta}}_k'\boldsymbol{x}_t)^2}{\sum_{t=2}^{T} p^*_{kt}}.$$

E Derivation of the prediction step in the Kalman filter

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{f}_t|\mathcal{I}_{t-1}, S_t = k] &= \mathbb{E}[\boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t|\mathcal{I}_{t-1}, S_t = k] \\
&= \boldsymbol{A}_k\mathbb{E}[\boldsymbol{f}_{t-1}|\mathcal{I}_{t-1}, S_t = k] + \mathbb{E}[\boldsymbol{\eta}_t|\mathcal{I}_{t-1}, S_t = k] \\
&= \boldsymbol{A}_k\hat{\boldsymbol{f}}_{t-1|t-1}, \text{ and} \\
\mathbb{V}[\boldsymbol{f}_t|\mathcal{I}_{t-1}, S_t = k] &= \mathbb{V}[\boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1} + \boldsymbol{\eta}_t|\mathcal{I}_{t-1}, S_t = k] \\
&= \boldsymbol{A}_k\mathbb{V}[\boldsymbol{f}_{t-1}|\mathcal{I}_{t-1}, S_t = k]\boldsymbol{A}_k' + \mathbb{V}[\boldsymbol{\eta}_t|\mathcal{I}_{t-1}, S_t = k] \\
&= \boldsymbol{A}_k\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}_k' + \boldsymbol{Q}_k, \text{ which implies that} \\
\boldsymbol{f}_t^k|\mathcal{I}_{t-1} &\sim \mathrm{N}(\boldsymbol{A}_k\hat{\boldsymbol{f}}_{t-1|t-1}, \boldsymbol{A}_k\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}_k' + \boldsymbol{Q}_k). \text{ Hence,} \\
\hat{\boldsymbol{f}}_{t|t-1}^k &= \boldsymbol{A}_k\hat{\boldsymbol{f}}_{t-1|t-1}, \\
\boldsymbol{P}_{t|t-1}^k &= \boldsymbol{A}_k\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}_k' + \boldsymbol{Q}_k.
\end{aligned}
$$

F Derivation of the update step in the Kalman filter

In the update step, we make use of the following lemma, described in Majumdar and Majumdar (2019):

$$\begin{aligned}
&\text{If } \begin{bmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \right), \text{ then} \\
&\boldsymbol{z}_2 | \boldsymbol{z}_1 \sim \mathrm{N}\Big(\boldsymbol{\mu}_2 + \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}(\boldsymbol{z}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Omega}_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{12} \Big).
\end{aligned} \tag{58}$$

Since $\boldsymbol{y}_t = \boldsymbol{c}_{S_t} + \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_t$, we know that

$$\begin{aligned}
&\boldsymbol{y}_t | \mathcal{I}_{t-1}, S_t = k \sim \mathrm{N}\Big(\boldsymbol{c}_k + \boldsymbol{\Lambda}_k \hat{\boldsymbol{f}}^k_{t|t-1}, \boldsymbol{\Lambda}_k \boldsymbol{P}^k_{t|t-1}\boldsymbol{\Lambda}'_k + \boldsymbol{R}_k \Big), \text{ which implies that} \\
&\begin{bmatrix} \boldsymbol{y}_t \\ \boldsymbol{f}_t \end{bmatrix} | \mathcal{I}_{t-1} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{c}_k + \boldsymbol{\Lambda}_k \hat{\boldsymbol{f}}^k_{t|t-1} \\ \hat{\boldsymbol{f}}^k_{t|t-1} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}_k \boldsymbol{P}^k_{t|t-1}\boldsymbol{\Lambda}'_k + \boldsymbol{R}_k & \boldsymbol{\Lambda}_k \boldsymbol{P}^k_{t|t-1} \\ \boldsymbol{P}^k_{t|t-1}\boldsymbol{\Lambda}'_k & \boldsymbol{P}^k_{t|t-1} \end{bmatrix} \right).
\end{aligned} \tag{59}$$

Hence, $\boldsymbol{f}_t | \boldsymbol{y}_t, \mathcal{I}_{t-1} \sim \mathrm{N}(\hat{\boldsymbol{f}}_{t|t}, \boldsymbol{P}_{t|t})$, with the following updates[13]:

$$\hat{\boldsymbol{f}}_{t|t} = \hat{\boldsymbol{f}}^{S_t}_{t|t-1} + \boldsymbol{P}^{S_t}_{t|t-1}\boldsymbol{\Lambda}'_{S_t}\Big(\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}^{S_t}_{t|t-1}\boldsymbol{\Lambda}'_{S_t} + \boldsymbol{R}_{S_t} \Big)^{-1}\Big(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\hat{\boldsymbol{f}}^{S_t}_{t|t-1} \Big), \tag{60}$$

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}^{S_t}_{t|t-1} - \boldsymbol{P}^{S_t}_{t|t-1}\boldsymbol{\Lambda}'_{S_t}\Big(\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}^{S_t}_{t|t-1}\boldsymbol{\Lambda}'_{S_t} + \boldsymbol{R}_{S_t} \Big)^{-1}\boldsymbol{\Lambda}_{S_t}\boldsymbol{P}^{S_t}_{t|t-1}. \tag{61}$$

[13]Since S_t is now known, we only update the factors and the corresponding uncertainties in the correct state. These updates are now simply denoted as $\hat{\boldsymbol{f}}_{t|t}$ and $\boldsymbol{P}_{t|t}$.

G Derivation of the inner expectation in the Kalman smoother

The joint distribution of $\boldsymbol{f}_t$ and $\boldsymbol{f}_{t+1}$ conditional on S_{t+1} and $\mathcal{I}_t$ is as follows:

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{f}_t \\ \boldsymbol{f}_{t+1} \end{bmatrix} |S_{t+1}, \mathcal{I}_t &\sim \mathrm{N}\left(\begin{bmatrix} \hat{\boldsymbol{f}}_{t|t} \\ \hat{\boldsymbol{f}}_{t+1|t} \end{bmatrix}, \begin{bmatrix} \boldsymbol{P}_{t|t} & \mathrm{Cov}[\boldsymbol{f}_t, \boldsymbol{f}_{t+1}] \\ \mathrm{Cov}[\boldsymbol{f}_{t+1}, \boldsymbol{f}_t] & \boldsymbol{P}_{t+1|t} \end{bmatrix} \right), \\
\text{with } \mathrm{Cov}[\boldsymbol{f}_t, \boldsymbol{f}_{t+1}] &= \mathbb{E}\Big[\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)\Big(\boldsymbol{f}_{t+1} - \hat{\boldsymbol{f}}_{t+1|t}\Big)'\Big] \\
&= \mathbb{E}\left[\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)\Big(\boldsymbol{A}_{S_{t+1}}\boldsymbol{f}_t + \boldsymbol{\eta}_{t+1} - \hat{\boldsymbol{f}}_{t+1|t}\Big)'\right] \\
&= \mathbb{E}\left[\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)\Big(\boldsymbol{A}_{S_{t+1}}\boldsymbol{f}_t - \boldsymbol{A}_{S_{t+1}}\hat{\boldsymbol{f}}_{t|t}\Big)'\right] \\
&= \mathbb{E}\left[\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)'\boldsymbol{A}'_{S_{t+1}}\right] \\
&= \mathbb{E}\left[\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)\Big(\boldsymbol{f}_t - \hat{\boldsymbol{f}}_{t|t}\Big)'\right]\boldsymbol{A}'_{S_{t+1}} \\
&= \boldsymbol{P}_{t|t}\boldsymbol{A}'_{S_{t+1}}.
\end{aligned}$$

Making use of the Normal lemma (58) again, we find the following:

$$\mathbb{E}[\boldsymbol{f}_t|\boldsymbol{f}_{t+1}, S_{t+1}, \mathcal{I}_t] = \hat{\boldsymbol{f}}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{A}'_{S_{t+1}}(\boldsymbol{P}^{S_t}_{t+1|t})^{-1}(\boldsymbol{f}_{t+1} - \hat{\boldsymbol{f}}^{S_t}_{t+1|t}).$$

H Derivation of parameter expressions in the M-step of the EM algorithm of the dynamic factor mixture model

We write the scalar in $\log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})$ as a trace and re-order the matrices within that trace to obtain the following:

$$
\begin{aligned}
\frac{\partial \log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})}{\partial \boldsymbol{A}_k} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{A}_k}\mathrm{Tr}\Big(\sum_{t=2}^{T}\boldsymbol{Q}_{S_t}^{-1}(\boldsymbol{f}_t - \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1})(\boldsymbol{f}_t - \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1})'\Big) = 0 \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{A}_k}\mathrm{Tr}\Big(\sum_{t=2}^{T}\boldsymbol{Q}_{S_t}^{-1}(\boldsymbol{f}_t\boldsymbol{f}_t' - \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1}\boldsymbol{f}_t' - \boldsymbol{f}_t\boldsymbol{f}_{t-1}'\boldsymbol{A}_{S_t}' \\
&+ \boldsymbol{A}_{S_t}\boldsymbol{f}_{t-1}\boldsymbol{f}_{t-1}'\boldsymbol{A}_{S_t}')\Big) \\
&= -\frac{1}{2}\sum_{t=2}^{T}\mathrm{I}[S_t = k]\Big(-(\boldsymbol{f}_{t-1}\boldsymbol{f}_t'\boldsymbol{Q}_k^{-1})' - \boldsymbol{Q}_k^{-1}\boldsymbol{f}_t\boldsymbol{f}_{t-1}' + (\boldsymbol{f}_{t-1}\boldsymbol{f}_{t-1}'\boldsymbol{A}_k'\boldsymbol{Q}_k^{-1})' \\
&+ \boldsymbol{Q}_k^{-1}\boldsymbol{A}_k\boldsymbol{f}_{t-1}\boldsymbol{f}_{t-1}'\Big) \\
&\implies \hat{\boldsymbol{A}}_k = \Big(\mathbb{E}\Big[\sum_{t=2}^{T}\mathrm{I}[S_t = k]\boldsymbol{f}_t\boldsymbol{f}_{t-1}'\Big]\Big)\Big(\mathbb{E}\Big[\sum_{t=2}^{T}\mathrm{I}[S_t = k]\boldsymbol{f}_{t-1}\boldsymbol{f}_{t-1}'\Big]\Big)^{-1} \qquad (62) \\
\frac{\partial \log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\Lambda}_k} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Lambda}_k}\mathrm{Tr}\Big(\sum_{t=1}^{T}\boldsymbol{R}_{S_t}^{-1}(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t)(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t)'\Big) = 0 \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Lambda}_k}\mathrm{Tr}\Big(\sum_{t=1}^{T}\boldsymbol{R}_{S_t}^{-1}((\boldsymbol{y}_t - \boldsymbol{c}_{S_t})(\boldsymbol{y}_t - \boldsymbol{c}_{S_t})' - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t(\boldsymbol{y}_t - \boldsymbol{c}_{S_t})' \\
&- (\boldsymbol{y}_t - \boldsymbol{c}_{S_t})\boldsymbol{f}_t'\boldsymbol{\Lambda}_{S_t}' + \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{\Lambda}_{S_t}')\Big) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathrm{I}[S_t = k]\Big(-(\boldsymbol{f}_t(\boldsymbol{y}_t - \boldsymbol{c}_k)'\boldsymbol{R}_k^{-1})' - \boldsymbol{R}_k^{-1}(\boldsymbol{y}_t - \boldsymbol{c}_k)\boldsymbol{f}_t' \\
&+ (\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{\Lambda}_k'\boldsymbol{R}_k^{-1})' + \boldsymbol{R}_k^{-1}\boldsymbol{\Lambda}_k\boldsymbol{f}_t\boldsymbol{f}_t'\Big) \\
&\implies \hat{\boldsymbol{\Lambda}}_k = \Big(\mathbb{E}\Big[\sum_{t=1}^{T}\mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)\boldsymbol{f}_t'\Big]\Big)\Big(\mathbb{E}\Big[\sum_{t=1}^{T}\mathrm{I}[S_t = k]\boldsymbol{f}_t\boldsymbol{f}_t'\Big]\Big)^{-1} \qquad (63) \\
\frac{\partial \log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})}{\partial \boldsymbol{c}_k} &= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{c}_k}\mathrm{Tr}\Big(\sum_{t=1}^{T}\boldsymbol{R}_{S_t}^{-1}(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t)(\boldsymbol{y}_t - \boldsymbol{c}_{S_t} - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t)'\Big) = 0 \\
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{c}_k}\mathrm{Tr}\Big(\sum_{t=1}^{T}\boldsymbol{R}_{S_t}^{-1}((\boldsymbol{y}_t - \boldsymbol{c}_{S_t})(\boldsymbol{y}_t - \boldsymbol{c}_{S_t})' - \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t(\boldsymbol{y}_t - \boldsymbol{c}_{S_t})' \\
&- (\boldsymbol{y}_t - \boldsymbol{c}_{S_t})\boldsymbol{f}_t'\boldsymbol{\Lambda}_{S_t}' + \boldsymbol{\Lambda}_{S_t}\boldsymbol{f}_t\boldsymbol{f}_t'\boldsymbol{\Lambda}_{S_t}')\Big) \\
&= -\frac{1}{2}\sum_{t=1}^{T}\mathrm{I}[S_t = k]\Big(-(\boldsymbol{y}_t'\boldsymbol{R}_k^{-1})' - \boldsymbol{R}_k^{-1}\boldsymbol{y}_t + (\boldsymbol{c}_k'\boldsymbol{R}_k^{-1})' + \boldsymbol{R}_k^{-1}\boldsymbol{c}_k + \boldsymbol{R}_k^{-1}\boldsymbol{\Lambda}_k\boldsymbol{f}_t + (\boldsymbol{f}_t'\boldsymbol{\Lambda}_k'\boldsymbol{R}_k^{-1})'\Big) \\
&\implies \hat{\boldsymbol{c}}_k = \frac{\mathbb{E}[\sum_{t=1}^{T}\mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{\Lambda}}_k\boldsymbol{f}_t)]}{\mathbb{E}[\sum_{t=1}^{T}\mathrm{I}[S_t = k]]} \qquad (64)
\end{aligned}
$$

Since the covariance matrices are symmetric, we can write

$$
\begin{aligned}
\frac{\partial \log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})}{\partial \boldsymbol{R}_k} &= \frac{1}{2}\sum_{t=1}^{T}\mathrm{I}[S_t = k]\boldsymbol{R}_k - \frac{1}{2}\sum_{t=1}^{T}\mathrm{I}[S_t = k](\boldsymbol{y}_t - \boldsymbol{c}_k - \boldsymbol{\Lambda}_k\boldsymbol{f}_t)(\boldsymbol{y}_t - \boldsymbol{c}_k - \boldsymbol{\Lambda}_k\boldsymbol{f}_t)' = 0 \\
&\implies \hat{\boldsymbol{R}}_k = \frac{\mathbb{E}[\sum_{t=1}^{T}\mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k - \hat{\boldsymbol{\Lambda}}_k\boldsymbol{f}_t)(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k - \hat{\boldsymbol{\Lambda}}_k\boldsymbol{f}_t)']}{\mathbb{E}[\sum_{t=1}^{T}\mathrm{I}[S_t = k]]} \qquad (65) \\
\frac{\partial \log f(\boldsymbol{y}_{1:T}, \boldsymbol{f}_{0:T}|\mathcal{I}_T, \boldsymbol{\theta})}{\partial \boldsymbol{Q}_k} &= \frac{1}{2}\sum_{t=2}^{T}\mathrm{I}[S_t = k]\boldsymbol{Q}_k - \frac{1}{2}\sum_{t=2}^{T}\mathrm{I}[S_t = k](\boldsymbol{f}_t - \boldsymbol{A}_k\boldsymbol{f}_{t-1})(\boldsymbol{f}_t - \boldsymbol{A}_k\boldsymbol{f}_{t-1})' = 0 \\
&\implies \hat{\boldsymbol{Q}}_k = \frac{\mathbb{E}[\sum_{t=2}^{T}\mathrm{I}[S_t = k](\boldsymbol{f}_t - \hat{\boldsymbol{A}}_k\boldsymbol{f}_{t-1})(\boldsymbol{f}_t - \hat{\boldsymbol{A}}_k\boldsymbol{f}_{t-1})']}{\mathbb{E}[\sum_{t=2}^{T}\mathrm{I}[S_t = k]]} \qquad (66)
\end{aligned}
$$

To write out the expectations in the equations above, we make use of $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] + \mathrm{Cov}[X,Y]$ and we write

$$\mathbb{E}[\boldsymbol{f}_t|\mathcal{I}_T] = \hat{\boldsymbol{f}}_{t|T}, \tag{67}$$

$$\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}_t'|\mathcal{I}_T] = \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t|T}' + \boldsymbol{P}_{t|T}, \tag{68}$$

$$\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}_{t-1}'|\mathcal{I}_T] = \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T}' + \boldsymbol{P}_{t,t-1|T}, \tag{69}$$

of which everything but $\boldsymbol{P}_{t,t-1|T}$ is given in the Kalman smoother. In Appendix I we show that $\boldsymbol{P}_{t,t-1|T} = \boldsymbol{P}_{t|T}(\boldsymbol{P}_{t|t-1}^{S_t})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1}$. Furthermore, we plug in the expectations in Equations 62-66 to get the following expressions[14]:

$$\hat{\boldsymbol{A}}_k = \Big(\sum_{t=2}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T}' + \boldsymbol{P}_{t,t-1|T})\Big) \times \Big(\sum_{t=2}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}_{t-1|T} + \boldsymbol{P}_{t-1|T}')\Big)^{-1}, \tag{70}$$

$$\hat{\boldsymbol{\Lambda}}_k = \Big(\sum_{t=1}^{T} \mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)\hat{\boldsymbol{f}}_{t|T}'\Big)\Big(\sum_{t=1}^{T} \mathrm{I}[S_t = k](\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t|T}' + \boldsymbol{P}_{t|T})\Big)^{-1}, \tag{71}$$

$$\hat{\boldsymbol{c}}_k = \frac{\sum_{t=1}^{T} \mathrm{I}[S_t = k](\boldsymbol{y}_t - \hat{\boldsymbol{\Lambda}}_k\hat{\boldsymbol{f}}_{t|T})}{\sum_{t=1}^{T} \mathrm{I}[S_t = k]}, \tag{72}$$

$$\begin{aligned}\hat{\boldsymbol{R}}_k = \frac{1}{\sum_{t=1}^{T} \mathrm{I}[S_t = k]} \sum_{t=1}^{T} \mathrm{I}[S_t = k]\Big(&(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)' - \hat{\boldsymbol{\Lambda}}_k\hat{\boldsymbol{f}}_{t|T}(\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)' \\ &- (\boldsymbol{y}_t - \hat{\boldsymbol{c}}_k)\hat{\boldsymbol{f}}_{t|T}'\hat{\boldsymbol{\Lambda}}_k' + \hat{\boldsymbol{\Lambda}}_k(\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t|T}' + \boldsymbol{P}_{t|T})\hat{\boldsymbol{\Lambda}}_k'\Big),\end{aligned} \tag{73}$$

$$\begin{aligned}\hat{\boldsymbol{Q}}_k = \frac{1}{\sum_{t=2}^{T} \mathrm{I}[S_t = k]} \sum_{t=2}^{T} \mathrm{I}[S_t = k]\Big(&\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t|T}' + \boldsymbol{P}_{t|T} - \hat{\boldsymbol{A}}_k(\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}_{t|T}' + \boldsymbol{P}_{t,t-1|T}') \\ &- (\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T}' + \boldsymbol{P}_{t,t-1|T})\hat{\boldsymbol{A}}_k' + \hat{\boldsymbol{A}}_k(\hat{\boldsymbol{f}}_{t-1|T}\hat{\boldsymbol{f}}_{t-1|T}' + \boldsymbol{P}_{t-1|T})\hat{\boldsymbol{A}}_k'\Big).\end{aligned} \tag{74}$$

[14] Since S_t is known for each time t, we have that $\mathbb{E}[\mathrm{I}[S_t = k]] = \mathrm{I}[S_t = k]$.

I Derivation of $\boldsymbol{P}_{t,t-1|T}$ in the E-step of the EM algorithm of the dynamic factor mixture model

To calculate $\boldsymbol{P}_{t,t-1|T}$, we first calculate $\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}'_{t-1}|\mathcal{I}_T]$ via the law of iterated expectations and then subtract $\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t-1|T}$ from it.

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}'_{t-1}|\mathcal{I}_T] &= \mathbb{E}[\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}'_{t-1}|\boldsymbol{f}_t,\mathcal{I}_T]|\mathcal{I}_T] \\
&= \mathbb{E}[\mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}'_{t-1}|\boldsymbol{f}_t,\mathcal{I}_{t-1}]|\mathcal{I}_T] \\
&= \mathbb{E}[\boldsymbol{f}_t\mathbb{E}[\boldsymbol{f}'_{t-1}|\boldsymbol{f}_t,\mathcal{I}_{t-1}]|\mathcal{I}_T] \\
&= \mathbb{E}[\big(\boldsymbol{f}_t\hat{\boldsymbol{f}}'_{t-1|t-1} + \boldsymbol{f}_t(\boldsymbol{f}'_t - \hat{\boldsymbol{f}}^{S_t}_{t|t-1}{}')(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1}\big)|\mathcal{I}_T] \\
&= \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t-1|t-1} + (\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T} - \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}^{S_t}_{t|t-1}{}')(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1} \\
&= \hat{\boldsymbol{f}}_{t|T}(\hat{\boldsymbol{f}}_{t-1|t-1} - \boldsymbol{P}_{t-1|t-1}\boldsymbol{A}'_{S_t}(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\hat{\boldsymbol{f}}^{S_t}_{t|t-1})' + (\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T})(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1} \\
&= \hat{\boldsymbol{f}}_{t|T}(\hat{\boldsymbol{f}}_{t-1|T} - \boldsymbol{P}_{t-1|t-1}\boldsymbol{A}'_{S_t}(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\hat{\boldsymbol{f}}_{t|T})' + (\hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}'_{t|T} + \boldsymbol{P}_{t|T})(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1} \\
&= \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T} + \boldsymbol{P}_{t|T}(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1}.
\end{aligned}
$$

Hence, $\boldsymbol{P}_{t,t-1|T} = \mathbb{E}[\boldsymbol{f}_t\boldsymbol{f}'_{t-1}|\mathcal{I}_T] - \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T} = \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T} + \boldsymbol{P}_{t|T}(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1} - \hat{\boldsymbol{f}}_{t|T}\hat{\boldsymbol{f}}_{t-1|T}$ $= \boldsymbol{P}_{t|T}(\boldsymbol{P}^{S_t}_{t|t-1})^{-1}\boldsymbol{A}_{S_t}\boldsymbol{P}_{t-1|t-1}$.

J Elaboration on the parameters in the simulation study

As stated in Section 3.1, we have high positive parameters in the first state, and intermediate parameters in the second state in the first model. Furthermore, we have intermediate parameters in the first state and low parameters in the second state in the second model. With high parameters we mean VAR matrices where the diagonal elements are random numbers between 0 and 0.5, and that each row has one off-diagonal element between 0.6 and 0.8, which is also randomly chosen, and 0 on all other entries[15]. The intermediate parameters are the same but then the randomly chosen off-diagonal is between 0.4 and 0.6. Lastly, the low parameters have an off-diagonal between 0.2 and 0.4.

[15]The random numbers can be reproduced using Matlab (MathWorks (2021)) with the default random seed.

K Scree plots of the Principal Component Analysis (PCA)

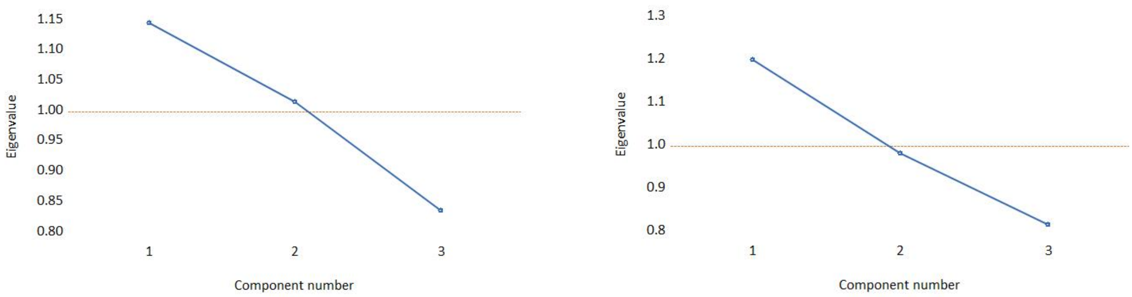

Figure 3: Scree plot of the PCA of all variables for Canada **(Left)**, and France **(Right)**.

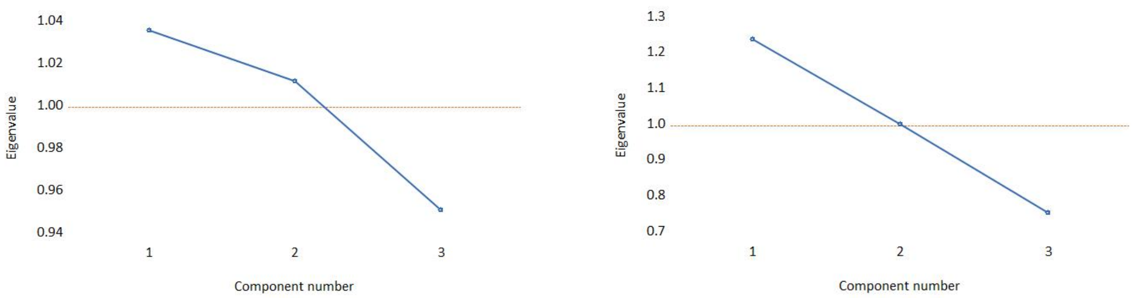

Figure 4: Scree plot of the PCA of all variables for Germany **(Left)**, and Italy **(Right)**.

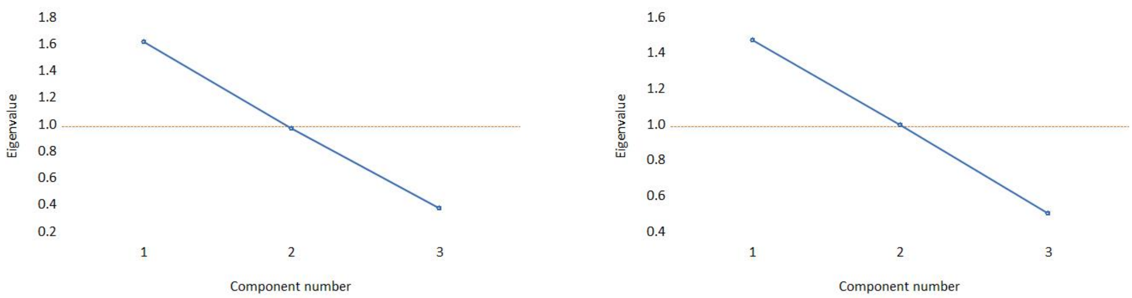

Figure 5: Scree plot of the PCA of all variables for Japan **(Left)**, and the UK **(Right)**.

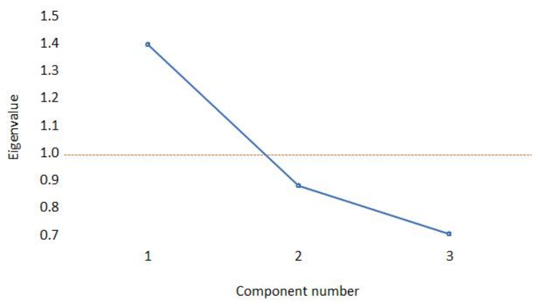

Figure 6: Scree plot of the PCA of all variables for the USA.

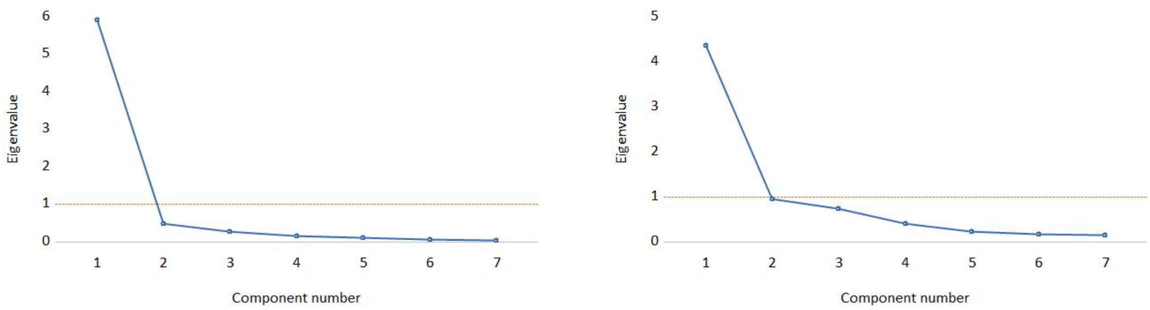

Figure 7: Scree plot of the PCA of the GDP growth for all countries **(Left)**, and the inflation rate for all countries **(Right)**.

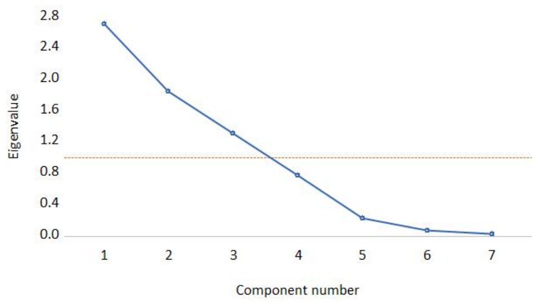

Figure 8: Scree plot of the PCA of the unemployment rate for all countries.

L Short description of the programming code

The zip-file containing the code consists of nineteen codes:

1. 'likelihoodfunc' - Calculates the likelihood that is needed in Equation 3 in the Hamilton filter.
2. 'Hamilton_filter' - Performs the Hamilton filter for the Markov switching model.
3. 'Hamilton_smoother' - Performs the Kim smoother for the Markov switching model.
4. 'EM_MS' - Runs one iteration of the EM algorithm of the Markov switching model.
5. 'Main_MS' - Estimates the Markov switching model and makes forecasts.
6. 'Kalman_filter' - Performs the Kalman filter for the dynamic factor model with constant parameters.
7. 'Kalman_smoother' - Performs the Kalman smoother for the dynamic factor model with constant parameters.
8. 'EM_DFM' - Runs one iteration of the EM algorithm of the dynamic factor model with constant parameters.
9. 'Main_DFM' - Estimates the dynamic factor model with constant parameters and makes forecasts.
10. 'Kalman_filter2' - Performs the Kalman filter for the dynamic factor mixture model.
11. 'Kalman_smoother2' - Performs the Kalman smoother for the dynamic factor mixture model.
12. 'EM_DFMM' - Runs one iteration of the EM algorithm of the dynamic factor mixture model.
13. 'Main_DFMM' - Estimates the dynamic factor mixture model and makes forecasts.
14. 'Kalman_filter3" - Performs the Kim filter for the Markov switching dynamic factor model.
15. 'Kalman_smoother3' - Performs the Kim smoother for the Markov switching dynamic factor model.
16. 'EM_MSDFM' - Runs one iteration of the EM algorithm of the Markov switching dynamic factor model.

17. 'Main_MSDFM' - Estimates the Markov switching dynamic factor model and makes forecasts.
18. 'Simulation_study' - Runs the simulation study.
19. 'Data' - Runs the application.