Markov-Switching Three Pass Regression Filter using a Regime-choosing diagnostic

Name: Jaïr Brandhorst
Student ID number: 478500

Bachelor Thesis: Econometrie and Operationele Research

Supervisor: Erik Kole
Second Assessor: Bram van Os

Date final version: 05-07-2020

Erasmus University Rotterdam
Erasmus School of Economics

Abstract

This paper aims to improve exchange rate forecasting by extending the MS-3PRF method proposed by Guérin et al. (2020) with a diagnostic that determines the appropriate amount of regimes for parameter estimation based on cycle synchronization measures described by Harding and Pagan (2006). The MS-3PRF with the diagnostic and MS-3PRF without the diagnostic are then compared to one another based on forecasting performance to determine whether the diagnostic shows any improvements in forecasting ability over the MS-3PRF. It was found that the diagnostic does not significantly improve the forecasting performance of the MS-3PRF. This could be attributed to the fact that the measure that was used to determine the amount of regimes, was not appropriate.

Contents

1 Introduction

Being able to make reliable forecasts of exchange rate time series has become increasingly important over the years, as indicated by the large amount of literature covering all the different methods and techniques to forecast exchange rates. Governments and Central Banks for example need reliable forecasts of exchange rates to be able to set their economic and monetary policies. On the other hand, consumers and pension funds that want to invest in currency need reliable forecasts of the exchange rates. However, exchange rates are notoriously hard to forecast. Meese and Rogoff (1983) for example state that the random walk without drift beats all, at that time available, forecasting methods for smaller forecast horizons. Forecasting exchange rates solely based on the random walk without drift is obviously not the most sophisticated method and it might therefore be possible to find better methods. Furthermore, Dennison (2018) argues that it is important for consumers and companies to increase their international investing. Exchange rate investing is one of the possibilities when investing internationally, which is why improving the existing methods and techniques is important.

Much of the existing literature regarding time series forecasting consists of research that focuses on methods where the parameters are assumed to be constant over time, however this might not be a realistic assumption as stated by Terasvirta and Anderson (1992), Beaudry and Koop (1993), Stock and Watson (1996) and Aastveit et al. (2017) among others. Stock and Watson (1996) even empirically found instability in exchange rate time series. Furthermore, Aastveit et al. (2017) found evidence of parameter instability following the 2008 financial crisis. Also, Koop and Potter (2001) recommend considering nonlinear models, specifically time-varying parameter models, when looking at time series. To deal with the instability of the parameters in exchange rate time series Guérin et al. (2020) proposed the Markov-Switching Three-Pass Regression Filter (MS-3PRF). This method allows time-varying parameters by using Markov-Switching parameters. Markov-Switching parameters are parameters that have different values for different states of the economy, called regimes. If the economy is in a certain state then the parameters are estimated according to the properties of that state. However, the state of the economy is unobservable, so methods using Markov-Switching parameters need to make predictions about the state of the economy. Which can be with the filters and smoothers proposed by Hamilton (1989) and Kim (1994).

The assumption here is that the economy changes from regime over time or following some big event. Therefore, some parts of the data need to be estimated according to the properties of one regime while other parts of the data need to be estimated according to another regime. State interpretations that are often used are a crisis state and a regular state or a state before some big event and a state after some big event.

The MS-3PRF also makes use of the 3PRF estimator proposed by Kelly and Pruitt (2015). Exchange rate time series often face the problem that the amount of predictors nears the same order of magnitude as the amount of observations, leading to inaccurate and inconsistent estimations. In general, methods making use of factor models are much better equipped to deal with these so-called dimensionality issues than exact methods, like for example Ordinary Least Squares (OLS). The 3PRF estimator is an example of such a method using factors to estimate the target variable instead of directly using the observations of the predictors.

Both of these features make the MS-3PRF a potentially valuable method for forecasting exchange rate time series. As exchange rate time series often have dimensionality issues and are shown to possibly benefit from being forecast using Markov-Switching parameters.

Guérin et al. (2020) show that their MS-3PRF, based on forecasting performance, seems to be a pretty good alternative to rivalling methods using factor modelling, when forecasting exchange rate time series. However, the MS-3PRF they used can still be improved upon, as they just laid the basic framework for the MS-3PRF. One of the aspects that can for example be improved upon is the amount of regimes that are used to estimate the Markov-Switching parameters. In their paper Guérin et al. (2020) use the MS-3PRF with a fixed amount of regimes, namely two regimes. However, this might not be appropriate as Psaradakis and Spagnolo (2003) and Garcia (1998) among others devised ways to determine the appropriate amount of regimes and Mussa et al. (2000) argued that exchange rate behavior seems to vary a lot over time. This indicates that there might be a need for methods that incorporate Markov-Switching with more than two regimes.

The aim of this paper is therefore to build upon the MS-3PRF by including a diagnostic that evaluates different amounts of regimes for every forecast and picks the amount of regimes that is most appropriate. The research question of this paper thus becomes: Can the MS-3PRF including a diagnostic which determines the appropriate amount of regimes beat the MS-3PRF based on forecasting performance for exchange rate time series?

The diagnostic is based on the cycle synchronization measures proposed by Harding and Pagan (2006). The idea of the diagnostic is that when the underlying cycles within the exchange rate time series that are used as predictors are not synchronized, additional regimes are needed to accurately estimate the target variable. When the cycles are synchronized then using two regimes is enough. The synchronization measures are the difference in mean of the time series and the correlation between the time series. The correlation matrix holds the most information and will thus be the basis of the decision making of the diagnostic. The Frobenius norm is finally used to calculate the distance of the distance of the correlation matrix to a perfect correlation matrix to measure the extent of synchronization. The diagnostic then picks the amount of regimes leading to the highest Frobenius norm. This diagnostic will be called the Diagnostic Markov Switching Three Pass Regression Filter (DMS-3PRF) or diagnostic throughout this paper.

The DMS-3PRF and MS-3PRF were compared to one another by first looking at the MS-3PRF using different amounts of regimes. This is done to see whether it is valuable at all to increase the number of regimes or if the MS-3PRF with two regimes performs just as well. Then the relation between the Frobenius norms and the squared forecast errors of the forecasts was determined to see if a higher Frobenius norm actually leads to lower squared forecast errors. Then some summary statistics of the Frobenius norms and amounts of regimes used within the DMS-3PRF are shown to give more context to the final conclusions. Ans lastly, the DMS-3PRF and MS-3PRF with two regimes are compared to one another directly based on forecasting performance. The data needed to obtain these results are obtained from databases from the International Financial Statistics of the International Monetary Fund (IMF) and from the St.Louis FED.

The results showed that the diagnostic was not able to significantly improve forecasting performance for the Canadian dollar, euro, British pound and Japanese yen compared to the MS-3PRF with two regimes. The results of the two methods gave extremely similar forecasts, while the results of the MS-3PRF with different amounts of regimes indicated that increasing the amount of regimes could possibly lead to better forecasts. The extreme similarity can be attributed to the fact that the Frobenius norm might not be an appropriate measure to base the decision of the amount of regimes on, as indicated by the results. The Frobenius norms were shown to have no significant relation with the squared forecast errors of the forecasts. Also the summary statistics of the Frobenius norms further justified the prior statement. Finally, the summary statistics also show that the diagnostic rarely chooses amounts of regimes larger than three, further indicating that the Frobenius norm might be higher on average for smaller amounts of regimes, while these smaller amounts of regimes do not necessarily lead to better forecasts.

The paper is structured as follows: Section 2 contains a review of the literature regarding the methods used in this paper. In section 3 the used methods are explained. The data is discussed in section 4. Then, section 5 contains the obtained results. Finally, section 6 concludes the paper. Section 7 contains the Appendix with some additional explanations.

2 Literature

2.1 Factor Models

As explained before, the aim of this paper is to improve exchange rate forecasting. A straightforward way to do this would be to use Ordinary Least Squares (OLS). Huber et al. (1973), however, show that when the amount of predictors is in the same order of magnitude as the number of time series observations the OLS estimates do not perform well. As exchange rate time series often are daily or monthly time series, the amount of time series observations can be quite low. The euro for example only has actual time series observations starting from 1999 (and estimated time series observations starting from 1979). On top of that, the amount of predictors is often quite high, as generally a lot of different exchange rates are used to forecast another exchange rate. Guérin et al. (2020) for example make use of 26 different exchange rates to forecast their target exchange rate. It therefore seems quite clear that OLS is not appropriate when forecasting exchange rates.

An alternative that could be appropriate is then the use of factor models. Factor models do not have the same issues with dimensionality that the OLS estimator has. The idea behind factor models is to assume both the target variable and predictors to be dependent on a set of unobservable factors that determine most of their variation. The set of factors on which the target variable is dependent is then assumed to be included in the set of factors on which the predictors are based. Methods using factor models therefore try to accurately extract estimates of the factors from the predictors and then use these factor estimates to compute forecasts for the target variable.

There is quite an extensive amount of literature covering different methods that make use of these factor models. Examples include Principal Component Analysis (PCA), formally proposed by Hotelling (1933), and different variations of PCA, like Principal Component Least Angle Regression (PC-LARS) and Targeted Principal Components Analysis (TPCA), both explained in more detail in the online appendix of Guérin et al. (2020). Two other examples are the Linear Three Pass Regression Filter (3PRF), proposed by Kelly and Pruitt (2015), and the Markov-Switching Three Pass Regression Filter (MS-3PRF), an extension of the 3PRF, proposed by Guérin et al. (2020). For a more complete picture of all the different methods using factor models I refer to Basilevsky (2009).

Factor models have also been used before in literature specifically regarding exchange rate forecasting. The work of Aguilar and West (2000) and Engel et al. (2015) among others, shows that factor models can be helpful when forecasting exchange rates.

2.2 Markov-Switching

A common assumption made in literature, explained by Goldfeld and Quandt (1973) is that exchange rate parameters have a finite and small amount of states, which are called regimes. The parameters vary considerably among these regimes, which means that, ideally, for every regime a different model should be used to estimate the parameters. Goldfeld and Quandt (1973) further explain that in time series analysis regimes can be interpreted in different ways, for example as different states in the business cycle or as states before and after a structural break. If it is possible to directly observe when a regime is going to switch then a different model could be used, until another regime-switch happens, to estimate the economic parameters. However, it is not possible to let a model switch based on these regimes as they are unobservable.

Therefore, Hamilton (1989) proposed the idea use so-called filters and smoothers to find estimates of these regimes based on previous observations. Hamilton (1990) proposed the Hamilton-filter which is used to determine forecast and inferred state probabilities, Kim (1994) proposed the Kim-filter which is used to determine smoothed probabilities. Important to note is that the Hamilton and Kim filters can only compute the probabilities at different points in time of the economy being in a certain regime. This means that the Hamilton and Kim filters cannot determine whether the economy is in a certain being, they can only give the probability of the economy being in a certain regime. This idea is commonly referred to as a regime-switching or Markov-switching model.

Markov-switching models are widely used in literature for all kinds of time series related research. Examples going from the valuation of American option prices, by Buffington and Elliott (2002), to crude oil futures prices, by Fong and See (2002). Markov-Switching models have seen quite a few attempts at improvement and extensions in more recent years and they have also been part of extensions within other methods.

Pelletier (2006), for example, used Markov-Switching to create a new model for the correlation between time series and Liu et al. (2011) used Markov-Switching within incorporated in DSGE models.

The Markov-Switching model has also been used for forecasting exchange rates. Klaassen (2002) for example managed to significantly improve exchange rate forecasts by making use of Markov-Switching. Other recent literature showing that Markov-Switching can be a valuable tool in the forecasting of exchange rates include the work of Bollen et al. (2000), Frömmel et al. (2005) and Wilfling (2009).

2.3 Cycles

When taking a look at time series, often certain patterns can be spotted really quickly. Think for example about ice cream sales decreasing during the winter and rising again during the summer. However, these patterns are not always as clearly visible. In many cases, just making use of the eye test is not enough to find the underlying patterns. There is thus a wide range of literature dedicated to methods for finding these patterns. Burns and Mitchell (1947) was the first that defined these patterns and called them business cycles. Building on this definition, Harding and Pagan (2002) describe cycles as "patterns in graphs of data on production, employment and prices". Following this definition it can be argued that exchange rate time series also have underlying cycles. The view of cycles that is used in this paper is the same as the view of Harding and Pagan (2006), who argue that the cycles have a set of turning points. At these turning points the cycle switches states. Following this definition there is then a latent unobservable variable, which represents the probability of being in that state of the economy.

The next step is to find out if these cycles can be used to improve exchange rate forecasting. Intuitively, cycles could offer some benefits if the cycles of different exchange rates are similar to one another. This similarity between a group of cycles is called the synchronization of that group. The synchronization of cycles can thus be described as the extent to which they move together or move opposite to one another. Harding and Pagan (2006) define multiple synchronization measures to quantify the synchronization between multiple cycles. They also proposed tests for significance of these synchronization measures. There is also some literature regarding the synchronization between cycles within exchange rate time series. McKinnon and Schnabl (2003) for example, looks at the synchronization between exchange rate cycles in East-Asian countries.

3 Methodology

The methods presented in this section are applied using MATLAB, which is computational software by MATLAB (2010). All the programs used to obtain the results following from the methods used in this section are self made. Furthermore, no additional packages or toolboxes were used. The Matlab code is contained in a accompanying zip-file.

3.1 Factor Models

Factor models are generally defined as follows:

$$x_t = \phi_0 + \boldsymbol{\phi}\boldsymbol{F}_t + \epsilon_t \tag{1}$$

where x_t is the set of predictors at time t. $\boldsymbol{F}_t$ is the set of common unobservable factors at time t. ϕ_0 and $\boldsymbol{\phi}$ are the intercept and the coefficients corresponding to the set of common unobservable factors respectively. Finally, ϵ_t is the idiosyncratic term at time t that captures the variation that is not captured by the common unobservable factors. A target variable, y, can now be forecast as follows:

$$y_{t+1} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t + \eta_{t+1} \tag{2}$$

Where, β_0 and $\boldsymbol{\beta}'$ are the intercept and the coefficients corresponding to the set of common unobservable factors respectively for the target variable and η_{t+1} is the idiosyncratic term at time t.

3.2 MS-3PRF

The linear 3PRF and MS-3PRF, proposed by Kelly and Pruitt (2015) and Guérin et al. (2020) respectively are examples of factor models. The MS-3PRF framework is defined as follows:

$$y_t = \beta_0(S_{yt}) + \boldsymbol{\beta}(S_{yt})\boldsymbol{f}_{t-1} + \eta_t, t = 1, ..., T \tag{3}$$

$$z_{j,t} = \lambda_{0,j}(S_{z_j,t}) + \boldsymbol{\lambda}_j(S_{z_j,t})\boldsymbol{f}_t + \omega_{j,t}, j = 1, ..., k_f \tag{4}$$

$$x_{i,t} = \phi_{0,i}(S_{x_i,t}) + \boldsymbol{\phi}_{f,i}(S_{x_i,t})\boldsymbol{f}_t + \boldsymbol{\phi}_{g,i}(S_{x_i,t})\boldsymbol{g}_t + \epsilon_{i,t}, i = 1, ...N \tag{5}$$

Where, $x_{i,t}$ is an element of $\boldsymbol{X}$, which is the NxT matrix of predictors, with N the amount of predictors and T the amount of time-series observations. y_t is an element of $\boldsymbol{y}$, which is the target variable that is going to be forecast. $z_{j,t}$ is an element of $\boldsymbol{Z}$, which is the k_fxT matrix of proxy variables, with k_f the amount common unobservable factors on which the predictors and the target variable are based. $\boldsymbol{f}$ is a k_fx1 vector of unobservable factors that fully drives $\boldsymbol{Z}$ and $\boldsymbol{y}$ and partially drives $\boldsymbol{Z}$, with corresponding coefficients $\boldsymbol{\beta}$, $\boldsymbol{\lambda}_j$ and $\boldsymbol{\phi}_{f,i}$. $\boldsymbol{g}$ is a k_gx1 vector of unobservable factors that also partially drives $\boldsymbol{X}$, but does not drive $\boldsymbol{y}$ and $\boldsymbol{Z}$, with corresponding coefficients $\boldsymbol{\phi}_{g,i}$. β_0, $\lambda_{0,j}$ and $\phi_{0,i}$ are the intercepts and η_t, $\omega_{j,t}$ and $\epsilon_{i,t}$ are the error terms. Finally, S_{yt}, $S_{z_j,t}$ and $S_{x_i,t}$ are M-state Markov Chains. Every Markov Chain has its own corresponding probability MxM matrix P_q, which is defined as follows:

$$P_q = \begin{bmatrix} p_{q,11} & p_{q,12} & \cdots & p_{q,1M} \\ p_{q,21} & p_{q,22} & \cdots & p_{q,1M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{q,M1} & p_{q,M2} & \cdots & p_{q,MM} \end{bmatrix} \text{ with } q = y_t,\ z_{j,t},\ x_{i,t}$$

This thus means that the intercepts $\beta_0(S_{yt})$, $\lambda_{0,j}(S_{z_j,t})$ and $\phi_{0,i}(S_{x_i,t})$ and the coefficients corresponding to $\boldsymbol{f}$ and $\boldsymbol{g}$, $\boldsymbol{\beta}(S_{yt})$, $\boldsymbol{\lambda}_j(S_{z_j,t})$, $\boldsymbol{\phi}_{f,i}(S_{x_i,t})$ and $\boldsymbol{\phi}_{g,i}(S_{x_i,t})$ are now allowed to switch between M regimes.

The extension of the linear-3PRF framework proposed by Guérin et al. (2020) is that the parameters within the linear-3PRF framework can switch between regimes following Markov Processes. This means that the linear 3PRF framework is the same as the framework depicted above with the exclusion of the Markov-Switching parameters S_{yt}, $S_{z_j,t}$ and $S_{x_i,t}$. Instead, the linear 3PRF framework assumes the parameters to be non-varying.

Important to take into account is that linear 3PRF and MS-3PRF rely on the assumption that the factors on which the target variable is based are a strict subset of the factors on which the predictors are based, or in other words the predictors and the target variable share a set of common factors which is a subset of the set of factors on which the predictors are dependent. This assumption is needed to make sure that the factor estimates extracted from the predictors can be used to compute estimates for the target variable.

3.2.1 Algorithm for MS-3PRF

The algorithm of the linear 3PRF is explained by Kelly and Pruitt (2015) in three steps: The first step involves running a time series regression of every single predictor on all the proxy variables, to acquire the so-called slope estimates. Then a cross-section regression is run of every predictor on coefficients constructed from the slope estimates for all observations, leading to factor estimates. Finally, a time series regression is run of the target variable on the factor estimates to obtain forecasts for the target variable. MS-3PRF extends linear 3PRF by allowing the parameters within step 1 and step 3 to switch between different regimes.

The algorithm for the MS-3PRF can thus be written as follows:

- Step 1: First N time-series regressions are run of $\boldsymbol{x}_i$ on $\boldsymbol{z}$ for $i = 1, \ldots, N$ to obtain slope estimates $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}$,

$$x_{i,t} = \phi_{0,i}(S_{x_i,t}) + \boldsymbol{z}'\boldsymbol{\phi}_{\boldsymbol{i}}(S_{x_i,t}) + \epsilon_{it} \quad t = 1, ..., T \tag{6}$$

  These slope estimates are used to compute the coefficients $\hat{\phi}_{A,it}$ and $\hat{\phi}_{B,it}$. The difference between these coefficients is explained in section 7.2 of the Appendix. Following the notation of Guérin et al. (2020) using $\hat{\phi}_{A,it}$ as the coefficients leads to the MS-3PRF and using $\hat{\phi}_{B,it}$ as the coefficients leads to the MSS-3PRF.

- Step 2: Then T cross-section regressions are run of $\boldsymbol{x}_t$ on $\hat{\boldsymbol{\phi}}_{q,it}$ for $i = 1, \ldots, N$, where q = A,B to obtain factor estimates $\hat{\boldsymbol{F}}_{\boldsymbol{t}}$,

$$x_{i,t} = \phi_{0,t} + \hat{\boldsymbol{\phi}}_q'\boldsymbol{F}_{\boldsymbol{t}} + \epsilon_{it} \quad i = 1, ..., N \tag{7}$$

- Step 3: Finally, a time series regression is run of y_t on estimated factors $\hat{\boldsymbol{F}}_{t-h}$, based on the forecast horizon, h, to obtain forecasts, $\hat{y}_t$

$$y_t = \beta_0(S_{yt}) + \hat{\boldsymbol{F}}' \boldsymbol{\beta}(S_{yt}) + \eta_t \tag{8}$$

As can be seen in the equation, this step again makes use of Markov-Switching parameters to estimate the intercept and the coefficients corresponding to the vector of factors. Following the notation of Guérin et al. (2020) this is called three pass MS(S)-3PRF. Another possibility is to not use Markov-Switching parameters, but instead use regular parameters, this is called first pass MS(S)-3PRF. This algorithm generates one h-period ahead forecast. The whole procedure has to be repeated for the all forecasts. The computation of the actual forecasts will follow in section 3.2.3.

Note that removing the Markov-Switching parameters leads to the original definition of the linear 3PRF as described by Kelly and Pruitt (2015).

3.2.2 Specification of the variables

What still remains to be explained in the MS-3PRF framework is firstly which variable to use as proxy variable. Also, it has to be specified how many factors are going to be used, thus it has to be specified what k_f is going to be. Kelly and Pruitt (2015) propose information criteria to determine the amount of factors, k_f, necessary. However, Guérin et al. (2020) explains that it could be more useful to determine the amount of factors empirically. This means that MS-3PRF is run a couple of times with different amounts of factors and find out what amount of factors lead to the best forecasts. Guérin et al. (2020) also explain that increasing the amount of factors could lead to reduced forecasting efficiency, while decreasing the amount of factors could introduce a bias. Which means the amount of factor has to be chosen carefully.

Kelly and Pruitt (2015) suggest, when there is just one $\boldsymbol{f}_t$ factor, to choose the proxy variable to be the target variable. When there are more factors they suggest using economic theory or using their automatic proxy-selection algorithm to choose the proxy variables. The algorithm is explained in detail in Kelly and Pruitt (2015). Since the research in this paper is aimed at the amount of regimes for the MS(S)-3PRF method, this paper will only feature MS-3PRF with just one factor and the proxy variable used is equal to the target variable.

3.2.3 Forecasts

Using the estimated coefficients from equation (8), $\hat{\beta}_0(S_{yt})$ and $\hat{\boldsymbol{\beta}}(S_{yt})$, forecasts for the target variable y_t can be made. As shown by Guérin et al. (2020), this can be done by doing the following computation:

$$\hat{y}_{t+h} = \sum_{j=1}^{M} \big(P(S_{yT+h} = j|\Omega_T)\hat{\beta}_0(S_{yT+h} = j) + P(S_{yT+h} = j|\Omega_T)\hat{\boldsymbol{\beta}}(S_{yT+h} = j)\hat{\boldsymbol{f}}_T \big), \tag{9}$$

Where, $P(S_{yT+h} = j|\Omega_T)$ is the predicted probability of being in regime j in period $T+h$ given all the available information up to time T, denoted by Ω_T.

Guérin et al. (2020) use the Mean Squared Forecast Error (MSFE) to compare the forecasts of the different methods to one another. The MSFE is calculated as follows:

$$\text{MSFE}_{model} = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2 \tag{10}$$

The lower the MSFE, the closer the sum of the differences between the actual values and the forecast values is to zero. Thus a method with a smaller MSFE has relatively better forecasts of the actual values than a method with a higher MSFE. Using the MSFE, the MS(S)-3PRF and the DMS(S)-3PRF can be compared to one another and to the Random Walk without drift. Rossi (2013) explains that this is the toughest benchmark to beat.

3.2.4 Random Walk without drift

The Random Walk without drift intuitively means that every next observation in the time series just randomly increases or decreases from the previous observation without any influence of other prior observations. Rossi (2013) describes this as follows:

$$E_t(y_{t+h} - y_t) = 0 \tag{11}$$

Where, E_t is the expectation of at time t, y_t is the value of the exchange rate at time t and h is the forecast horizon. This means that h-period ahead forecasts are just the same value as the final value of the estimation period. This is then done for all forecasts. As can be seen, this method does not take into account any extra information from other exchange rates.

3.3 EM-algorithm

Estimation of the parameters in equations 3, 4, 5 is done by performing maximum likelihood. However, when using Markov-Switching parameters this likelihood gets computationally demanding, which is why the Expectation-Maximization (EM) algorithm is used, proposed by Dempster et al. (1977). This algorithm does not have the same computational issues as regular maximum likelihood and is shown by Dempster et al. (1977) to be a consistent estimator.

This algorithm has two basic steps, an expectation step and a maximization step, which are repeated until a pre-defined stopping condition is reached. This algorithm is explained in more detail in the Appendix. The expectation step includes the computation of the forecast probabilities following from the Hamilton and Kim filters. The maximization step includes the maximization of the parameter vector to find a new improved parameter vector. The mathematics behind the expectation and maximization steps is further explained in the Appendix.

3.4 Synchronization

There are couple of different types of synchronization between two cycles (or possibly more than two). Strong perfect positive synchronization (SPPS) means two cycles, S_{y_1t} and S_{y_2t}, are exactly the same, meaning when $S_{y_1t} = 1$, $S_{y_2t} = 1$ as well. Strong perfect negative synchronization means the two specific cycles, S_{y_1t} and S_{y_2t}, are exactly opposite to one another, meaning when $S_{y_1t} = 1$, $S_{y_2t} = 0$. Finally, strongly non-synchronized (SNS) cycles are two specific cycles, S_{y_1t} and S_{y_2t}, that are independent of one another, which means that $S_{y_1t} = 1$ has no influence on whether S_{y_2t} has value one or zero.

Harding and Pagan (2006) then proposed three measures of synchronization. Two of these measures are based on the definition of SPPS and the third is based on the definition of SNS cycles. In this paper only the first two measures based on SPPS are used.

The first measure computes the distance between the expected values of the cycles, $\mu_{S_{y1}}$ and $\mu_{S_{y2}}$. In other words, it computes $\mu_{S_{y1}} - \mu_{S_{y2}}$. SPPS requires the distance to be zero. For multivariate cycles this means that the distance between every possible pair of cycles has to be measured. SPPS requires all entries of the matrix following from all distances to be equal to zero.

The second measure computes the correlation ρ_S between two cycles, where ρ_S is based on properties of both cycles S_{y_1t} and S_{y_2t}. ρ_S is calculated in the same way that the correlation between to vectors is normally calculated. SPPS requires the correlation to be one. For multivariate cycles this means that the correlation between every possible pair of cycles has to be measured. SPPS requires all entries of the matrix following from all correlations to be equal to one.

A value close to zero for the first measure and a value close to one (or minus one) for the second measure thus hints at a high level of synchronization. For multivariate cycles this is slightly more complicated as the correlation and expected value difference are not values but matrices, which means distance measures are not as intuitive. A metric like the Frobenius norm can be used to calculate the distance between the correlation matrix and a matrix of ones (or minus ones or a combination of both) and the distance between the expected value differences matrix and a matrix of zeros. The Frobenius norm is normally calculated as follows:

$$||A||_F = \sqrt{\Big(\sum_{p=1}^{P}\sum_{q=1}^{Q} a_{pq}^2\Big)} \tag{12}$$

Where, A is a PxQ matrix and a_{pq}^2 are the squared elements of A. The Frobenius norm of a matrix of all ones (or minus ones, or both) is equal the size of the matrix. Thus in this case the Frobenius norm of a correlation matrix with SPPS in the cycles is equal to the amount of predictors. An important note here is that the Frobenius norm will only be used to compute the distance between the correlation matrix and a matrix of ones (or minus ones or a combination of both). As the correlation matrix offers much more information about whether the cycles are actually synchronized or not.

Another possibility is to test the synchronization measures using the formal tests. In this paper the first option was chosen, since the cycles of all exchange rate time series are extremely unlikely to be totally synchronized. A test to find out whether they are actually totally synchronized is then unnecessary. For further explanation of these formal test, I refer to Harding and Pagan (2006).

3.5 Diagnostic

The idea of this paper is first to create a diagnostic that helps with forecasting exchange rates. This diagnostic first estimates the model using two regimes and then using three regimes. Then it determines the synchronization between the cycles for all the exchange rate series based on the synchronization measures and the Frobenius norm as a distance between the correlation and a matrix of all ones for both two and three regimes. If the Frobenius norm for two regimes is larger than that of two regimes, the diagnostic will estimate the model using four regimes and check what the Frobenius norm is. This process continues until the Frobenius norm of the model with an additional regime is smaller. In this way the diagnostic is used to determine whether to proceed to the next iteration with an additional regime or not.

With the help of the diagnostic the right amount of regimes should now be picked in the first step of the MS-3PRF for every forecast. This means that if a large Frobenius norm leads to a smaller forecast error, which we assume here since a large Frobenius norm corresponds with a correlation matrix closer to a prefect correlation matrix, that the diagnostic leads to forecasts with smaller forecast errors. The diagnostic, however, does have the practical flaw that for every forecast it estimates the N slope-estimates at least two times and even more when the diagnostic picks a larger amount or regimes. This can dramatically increase computation times.

The inclusion of the diagnostic within the algorithm can be written as follows:

- Step 1: First N time-series regressions are run of $\boldsymbol{x}_i$ on $\boldsymbol{z}$ for $i = 1, \ldots, N$ to obtain slope estimates $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}$,

$$x_{i,t} = \phi_{0,i}(S_{x_i,t}) + \boldsymbol{z}'\boldsymbol{\phi_i}(S_{x_i,t}) + \epsilon_{it} \quad t = 1, ..., T \tag{13}$$

  Now the synchronization measures and the corresponding Frobenius norm are calculated. The above regressions are repeated, but now allowing for three possible regimes. The synchronization measures and the Frobenius norms are calculated again. Both norms are compared to one another. If the Frobenius norm indicates that an additional regime is necessary then the above regressions are again repeated for an additional regime and the norms are again compared. If the Frobenius norm indicates that an additional regime is not necessary then the estimates of the regime with the highest Frobenius norm are used to move on to the computation of $\hat{\phi}_{A,it}$ and $\hat{\phi}_{B,it}$.

4 Data

The aim of this paper is forecasting exchange rates. In order to do that exchange rate data is needed. This data will be the same as the data used in Greenaway-McGrevy et al. (2018). The data is obtained from the International Financial Statistics of the International Monetary Fund (IMF). IMF has many different exchange rates databases. In this paper we use the databases where the exchange rates are based on the US dollar, as was also done by Guérin et al. (2020). However, where Guérin et al. (2020) used the database with the exchange rates as a monthly average of the daily exchange rates, the database with the end of period monthly exchange rates data will be used. This is because end of period data is more often used by investors to base their decisions on. Monthly average data does not give a good representation of the daily data.

Guérin et al. (2020) used a dataset running from January 1995 to December 2015. This dataset thus has 252 observations (as there are 252 months) for 26 exchange rates. Just like Guérin et al. (2020) the exchange rates from the following countries are included: Australia, Brazil, Canada, Chile, Columbia, the Czech Republic, the euro area, Hungary, Iceland, India, Israel, Japan, Republic of Korea, Mexico, Norway, New Zealand, the Philippines, Poland, Romania, Singapore, South Africa, Sweden, Switzerland, Taiwan, Turkey and the United Kingdom. Important to note is that the euro, the currency for the euro area, was only introduced in 1999. This means there are no actual observations available in the International Financial Statistics for the period 1995-1999 for the euro. However, the St.Louis FED does have a database that contains estimated exchange rate data from 1979-1999 for the euro, based on exchange rates of other currencies. I will be making use of this data set to complete any missing values in the data set obtained from the International Financial Statistics.

Table 1: This table shows the average, standard deviation, maximum and minimum for every exchange rate series based on the US dollar for the time period Jan 1995 - Dec 2015.

Countries	Average	Standard deviation	Maximum	Minimum
Australia	46.35	7.86	66.60	31.37
Brazil	2.03	0.69	3.89	0.84
Canada	1.236	0.71	3.01	0.04
Chile	31.47	2.28	35.07	25.42
Colombia	214.67	41.19	308.27	111.45
Czech Republic	90.62	25.07	136.82	59.94
Euro Area	0.84	0.13	1.17	0.63
Hungary	10.89	2.26	17.00	5.51
Iceland	1.26	0.20	1.60	0.96
India	25.70	6.70	41.13	14.92
Israel	536.08	87.24	745.21	373.59
Japan	7.64	1.10	10.78	5.95
Republic of Korea	1986.34	535.17	3246.44	846.67
Mexico	1099.25	165.73	1701.53	757.01
Norway	1.35	0.27	2.00	0.93
New Zealand	7.49	2.25	14.99	3.54
Philippines	44.09	8.50	56.34	24.62
Poland	6.82	1.04	9.36	5.05
Romania	107.65	14.75	144.66	76.77
Singapore	2.51	1.10	4.14	0.18
South Africa	0.61	0.05	0.71	0.48
Sweden	1.51	0.19	1.84	1.21
Switzerland	1.23	0.25	1.78	0.78
Taiwan	3.33	0.56	4.64	2.07
Turkey	3.93	0.47	4.95	2.96
United Kingdom	1.59	0.34	2.51	1.15

Table 1 shows some summary statistics for the exchange rate series data. These summary statistics give context to the nature of the exchange rate time series. Note that the summary statistics for the euro include the estimated end of period monthly exchange rates obtained from the St.Louis FED.

Before performing any of the methods, the data will first be transformed. As the log differences of the data will be used. A log difference series of the data, which will be called r here, is computed as follows:

$$r_t = \log(R_{t+1}) - \log(R_t), \quad t = 1, ..., (T-1) \tag{14}$$

Where, R_t is observation t in the regular exchange rate series and T is the total amount of observations in R_t. Obviously since the differences are computed here, the amount of observations in r_t is equal to $T - 1$.

Following Guérin et al. (2020), the exchange rates that are forecast are the Canadian dollar, the Japanese yen, the euro and the British pound. Furthermore, the estimation sample will be January 1995 to July 2006. This sample is used to compute the h-period ahead forecasts, where h is the forecast horizon. After having completed one forecast, the whole process is repeated and another forecast is computed. For this next forecast the estimation sample is expanded using the expanding window technique. This is done until all forecasts are computed. The forecast horizon is set to h = 1, 2, 3, 6, 9 and 12 months to compare results for different forecast horizons.

5 Results

In order to find out the value of the diagnostic, several steps have to be taken. Firstly, forecasts using MS-3PRF and MSS-3PRF with different amounts of allowed regimes are compared against one another based on their MSFE. By comparing the methods with different regimes it becomes clear whether increasing the amount of regimes would yield any increase in forecasting performance at all. Then, using the Frobenius norm and the squared forecast errors for every forecast a regression is run to see if, when the Frobenius norm increases, the squared forecast error decreases on average. If this relation is significant then this would point to the fact that a high Frobenius norm leads to improved forecasts. The use of the Frobenius norm as a metric for the diagnostic to base the decision of the amount of regimes on can then be justified. Finally, the diagnostic is compared to the MS-3PRF and MSS-3PRF with two regimes. By comparing these methods to one another it becomes clear whether the diagnostic increases forecasting performance.

5.1 Different amounts of regimes

First, the MS-3PRF and MSS-3PRF with two, three, four and five regimes were used to compute forecasts for the euro for six different forecast horizons, h = 1, 2, 3, 6, 9 and 12. For these forecasts, the MSFE was computed. The results are gathered in table 2.

Table 2: This table shows the MSFE ($\cdot 10^{-3}$) of all the performed methods. MS-3PRF uses coefficient A and MSS-3PRF uses coefficient B. The number in brackets stands for the amount of regimes used. Finally, RW is the Random Walk method. The best results per forecast horizon appear in bold. Significantly different results from MS(S)(2) based on Diebold Mariano test appear with a *.

EUR	H	MS(2)	MS(3)	MS(4)	MS(5)	RW
	1	0.661	0.651	0.651	**0.649**	0.892
	2	0.614	**0.598**	0.609	0.606	1.158
	3	0.619	0.620	**0.613**	**0.613**	1.059
	6	0.621	0.620	0.616	**0.602**	1.201
	9	0.622	0.625	**0.606**	0.610	1.359
	12	0.625	0.623	**0.620**	0.628	1.329
	H	MSS(2)	MSS(3)	MSS(4)	MSS(5)	RW
	1	0.642	**0.636**	0.646	0.647	0.892
	2	0.611	0.609	**0.607**	0.621	1.158
	3	**0.611**	0.614	0.621	0.614	1.059
	6	0.619	**0.613**	0.635	0.625	1.201
	9	0.612	**0.611**	0.618	0.625	1.359
	12	0.620	0.616	**0.613**	**0.613**	1.329

Table 2 shows that when forecasting the euro exchange rate series, there are no significant differences between using MS(S)-3PRF with different amounts of regimes. The differences in MSFE between the different regimes are quite small and it can also be seen that there seems to be no clear cut best amount of regimes. Furthermore, there seems to be no pattern among forecast horizons. In other words, the results do not indicate that a certain amount of regimes should always be used for a certain forecast horizon. Also, it can be seen that increasing the amount of regimes does not necessarily lead to better forecasting results.

5.2 The diagnostic

5.2.1 Relation Frobenius norm and squared forecast errors

Following the comparison of the MS-3PRF and MSS-3PRF methods allowing for different amounts of regimes, we will now look at whether the Frobenius norm is actually a good measure for determining what amount of regimes the diagnostic should pick. This can be done by looking at the relation between the Frobenius norm and the squared forecast errors following from the diagnostic. The relation can be looked at by means of regressing the Frobenius norms of all forecasts on the squared forecast errors of all forecasts. The coefficient corresponding to the Frobenius norms and its standard deviation are used to derive a t-statistic: $t = \frac{\text{Mean}(c)}{\text{stdev}(c)}$, where c stands for the coefficient corresponding to the Frobenius norm. The critical values are 2 and -2 for this t-statistic for a significance level of 5 percent. Table 3 shows the results of the regressions mentioned above for all six forecast horizons.

Table 3: This table shows the t-values of the regressions of the Frobenius norms on the squared forecast errors of all forecasts for the six forecast horizons H for the euro. Significant values are marked by *.

	EUR		JAP		CAN		GBR	
H	DMS	DMSS	DMS	DMSS	DMS	DMSS	DMS	DMSS
1	0.086	0.072	0.132	0.144	-0.050	0.067	0.445	0.433
2	0.072	0.077	0.123	0.141	0.090	0.006	0.436	0.441
3	0.099	0.084	0.134	0.127	-0.003	0.081	0.414	0.418
6	0.079	0.084	0.133	0.143	0.363	0.165	0.421	0.352
9	0.090	0.085	0.156	0.176	-0.006	0.155	0.478	0.466
12	0.090	0.088	0.149	0.136	-1.004	0.108	0.452	0.455

As is shown in table 3 none of the t-values are significant. All t-values are in fact quite far off from being significant. It can also be seen that the results do seem to differ among different exchange rate series, as for the British pound the values are somewhat higher than for the other exchange rate series. This difference however, is again not significant. Finally, there does not seem to be a difference between the methods in t-values.

5.2.2 Amount of regimes used in the diagnostic

It might be important to know how often the diagnostic chooses to increase the amount of regimes. As this gives more insight into whether the diagnostic chooses the same amount of regimes often or whether it chooses many different amounts of regimes. This could then give more information on whether the Frobenius norm is a good measure for the diagnostic to base its decision on. As was shown in table 2 the MSFE's of the forecasts of the methods using different regimes, remained close to one another. Possibly suggesting that increasing the amount of regimes for some forecasts might prove to be beneficial, while for others it might not be beneficial. A diagnostic picking the best amount of regimes for every forecast might therefore lead to an increase in forecasting performance. Note that the choosing of the amount of regimes happens before picking the set up of the coefficient constructed from the slope estimates and thus whether to use MS-3PRF or MSS-3PRF. It also happens before the forecast horizon is used. Therefore, the amount of regimes is the same for all forecast horizons for both MS-3PRF and MSS-3PRF. Table 4 shows the amount of forecasts using a certain amount of regimes.

Table 4: This table shows the amount of times a certain amount of regimes was used by the diagnostic for the Canadian dollar (CAN), British pound (GBR), Japanese yen (JAP) and the euro (EUR).

Currency	Amount of regimes			
	2	3	4	5
JAP	104	8	0	0
CAN	111	1	0	0
GBR	112	0	0	0
EUR	90	19	3	0

Table 4 shows that the diagnostic based on the Frobenius norm very rarely chooses to increase the regimes. Though this does vary quite a bit from currency to currency, as for the euro the diagnostic increases the amount of regimes 22 times, while for the British pound the diagnostic does not increase the amount of regimes once.

Table 5 shows the average Frobenius norm for the different amounts of regimes used.

Table 5: This table shows the average Frobenius norm for a certain amount of regimes calculated by the diagnostic for the Canadian dollar (CAN), British pound (GBR), Japanese yen (JAP) and the euro (EUR). Standard deviations appear in brackets behind the averages.

Currency	Amount of regimes			
	2	3	4	5
JAP	7.882 (0.604)	7.496 (0.335)	7.026 (0.234)	- (-)
CAN	7.959 (0.238)	7.132 (0.146)	6.905 (0.000)	- (-)
GBR	7.578 (0.373)	6.988 (0.325)	- (-)	- (-)
EUR	7.613 (0.524)	7.340 (0.360)	7.237 (0.112)	6.764 (0.151)

Table 5 shows that on average the Frobenius norm when using two regimes can be quite a bit higher than the Frobenius norm using three regimes. In general it seems that the higher the amount of regimes is, that the lower the value of the Frobenius norm and the standard deviation becomes. An important note here is that the diagnostic only calculates an additional regime when necessary. The results in table 5 following from the fourth and fifth regimes are therefore sometimes missing or are obtained with a very low amount of observations, which is why the results for these amounts of regimes might be inaccurate. Furthermore, for the Canadian dollar the diagnostic chose three regimes only once, which means the average of four regimes is based on one observation and the standard deviation is zero.

5.3 Synchronization

Finally, the MSFE of the forecasts for the different methods, MS-3PRF, MSS-3PRF, DMS-3PRF and DMSS-3PRF, to see whether the methods incorporating a diagnostic have a better forecasting performance. The

MS(S)-3PRF methods use two regimes as that is the method that the diagnostic is trying to improve upon. Table 6 shows the MSFE for the forecasts created from the MS(S)-3PRF and the DMS(S)-3PRF.

Table 6: This table shows the MSFE ($\cdot 10^{-3}$) of all the performed methods. MS-3PRF uses coefficient A and MSS-3PRF uses coefficient B. The number in brackets stands for the amount of regimes used. DMS(S)-3PRF is MS(S)-3PRF including the diagnostic. RW stands for Random Walk. The best results per forecast horizon appear in bold. Significantly different results from MS(S)(2) based on Diebold Mariano test for DMS(S) appear with a *.

EUR	H	MS(2)	DMS	MSS(2)	DMSS	RW
	1	0.661	0.661	0.642	**0.641**	0.892
	2	0.614	0.613	0.611	**0.611**	1.158
	3	0.619	0.619	**0.611**	0.611	1.059
	6	0.621	0.621	0.619	**0.619**	1.201
	9	0.622	0.622	**0.612**	0.613	1.359
	12	0.625	0.625	0.620	**0.620**	1.329
JAP	H	MS(2)	DMS	MSS(2)	DMSS	RW
	1	**0.872**	0.872	0.889	0.889	0.969
	2	0.851	0.850	**0.805**	0.806	1.404
	3	0.856	0.856	0.813	**0.813**	1.352
	6	0.845	0.845	0.823	**0.822**	1.584
	9	0.849	0.849	0.823	**0.823**	2.019
	12	0.809	0.809	**0.788**	0.788	1.745
CAN	H	MS(2)	DMS	MSS(2)	DMSS(2)	RW
	1	1.311	1.304	**1.258**	1.265	1.495
	2	1.241	**1.240**	1.259	1.262	2.207
	3	1.225	**1.225**	1.233	1.240	2.089
	6	1.110	**1.110**	1.236	1.234	2.089
	9	1.176	1.161	1.178	**1.172**	2.221
	12	1.129	1.128	**1.113**	1.118	2.317
GBR	H	MS(2)	DMS	MSS(2)	DMSS	RW
	1	1.003	1.003	**1.002**	**1.002**	1.304
	2	1.003	1.003	**1.002**	**1.002**	1.665
	3	1.004	1.004	**1.002**	**1.002**	1.587
	6	1.000	1.000	**0.974**	**0.974**	2.025
	9	1.002	1.002	**1.001**	**1.001**	2.094
	12	**1.001**	**1.001**	1.002	1.002	2.194

Table 6 shows that the MSFE of the diagnostic compared to the MS(S)-3PRF excluding the diagnostic are very close to another. So close in fact that the MSFE's rounded to the first three decimals are almost always equal. For the the British pound this is not surprising. As table 4 shows that for the British pound the amount of regimes used in forecasts by the diagnostic is always two regimes. This leads to the MS(S)-3PRF and DMS(S)-3PRF forecasts to be equal for almost every forecast and therefore leads to the MSFE being almost exactly equal. The fact that the results are so close for the euro and the Japanese yen is more surprising. Since the diagnostic did choose different regimes for a couple of forecasts for these currencies, although they still mostly chose two regimes. Also, the fact that the MSFE for the Canadian dollar are the furthest apart is also quite surprising since the diagnostic only once chose three regimes when forecasting for the Canadian dollar.

What can be seen in table 6 is that using the diagnostic does not guarantee that the MSFE of the resulting forecasts is smaller than the MSFE of the MS(S)-3PRF. This shows that even though the diagnostic might have a smaller MSFE for some forecast horizons, it also could have a higher MSFE for other forecast horizons.

Finally, the diagnostic does not seem to show any particularly large improvements for a certain forecast horizon. In other words, smaller forecast horizons do not seem to benefit anymore from the diagnostic than larger forecast horizons and vice versa.

6 Conclusion

6.1 Discussion of the results

In conclusion, compared to the MS(S)-3PRF, the diagnostic did not significantly improve on forecasting performance for the British pound, Canadian dollar, euro and Japanese yen. Even though the results for the MS(S)-3PRF using different amounts of regimes possibly indicated that some forecasts could be computed more accurately using MS-3PRF with more than two regimes.

The fact that the diagnostic did not lead to an improvement might be attributed to the fact that the Frobenius norm might not have been the best measure to base the diagnostics choosing of the regimes on. Since it was shown that the Frobenius norm did not have a negative relation with the squared forecast errors. This negative relation would have indicated that when the Frobenius norm increased, the squared forecast errors would have decreased. This could have justified the diagnostic choosing the appropriate amount of regimes based on the Frobenius norm. Also, it was shown that the diagnostic tends to lean heavily towards picking two regimes for the forecasts, which lead to the diagnostic almost seeing no improvements compared to the MS(S)-3PRF. The fact that the diagnostic likes to choose two regimes can be attributed to the fact that on average the Frobenius norm seemed to decrease with amount of regimes increasing. While, the results showing the differences between the MS(S)-3PRF with different regimes indicated that forecasts using MS(S)-3PRF with three and four regimes were just as good if not even better than the forecasts produced by MS(S)-3PRF with two regimes.

Finally, it was shown that even though the diagnostic did choose more than two regimes for some of the forecasts for the euro and the Japanese yen the MSFE's following from these results were still extremely close to the MSFE's of the MS(S)-3PRF for these currencies. This might be further indication that the Frobenius norm is not an appropriate measure for the diagnostic to base its decision on, in that the diagnostic did not choose the correct forecasts that needed additional regimes, but instead also chose a few forecasts to be estimated with more than two regimes while this was not necessary. However, it could also indicate that maybe the diagnostic did choose the correct forecasts that needed additional regimes, but the additional regimes did not increase the forecasting performance. However, this second explanation seems to be more shaky as the forecasts for the Canadian dollar differed the most. While the diagnostic picked three regimes only for one forecast for the Canadian dollar. In summary, based on the results obtained in this paper a definitive conclusion on the true effectiveness of the MS-3PRF method with diagnostic can not be stated. What can be concluded though is that the Frobenius norm following from the synchronization measures is not an appropriate measure to base the decision of additional regimes for forecasts on.

6.2 Further Research

Even though the results did not show any significant improvements, I still think that this method has potential. Since in this paper the diagnostic made use of a measure that turned out to be an inappropriate measure to base the decision of the amount of regimes for the forecasts on. However, as stated by Guérin et al. (2020) the MS(S)-3PRF performs reasonably well compared to rivalling methods and the Random Walk without drift method. Therefore, if a measure can be found that does have a significant relation with the squared forecast errors, then that measure could be used to improve upon a method that already performs well.

Other directions of research include looking at the amount of factors allowed, which is determined by the amount of proxy variables used. In this paper the amount of factors was kept at one, however Kelly and Pruitt (2015) already proposed an algorithm that can determine what proxy variables to use, which would also increase the amount of factors used. Furthermore, in this paper the N Markov-Switching time series regressions done in step 1 are calculated one by one apart from one another. However, it might also be possible to do all these regressions at once. It would be interesting to see if and how that would change the forecasting performance of the MS-3PRF. Finally, the data used in this paper is exchange rate time series data from all kinds of currencies all over the world. It would be interesting to look at the differences between forecasting performance for exchange rate time series for different countries and see if there are patterns for different types of currencies. As in this paper it can already be seen that the MSFE among the four different currencies could be quite different.

7 Appendix

7.1 EM-Algorithm

The EM-Algorithm has two basic steps that are constantly repeated. First the algorithm computes the expectations for a given parameter vector, then the log-likelihood function is maximized, which leads to a new parameter vector. This new parameter vector is then used to compute new expectations and this process repeats itself until a self-specified stopping criterion is met. Both the expectation step and the maximization step will be explained in detail in this part.

7.1.1 Expectation step

In the expectation step, three types of probabilities are calculated, the inference, forecast and smoothed probabilities. These probabilities are computed by making use of two recursions, which are called the Hamilton Filter and the Kim Filter, proposed respectively by Hamilton (1989) and Kim (1994). Forecast state probabilities use all the information available until observation $t-1$ to compute for observation t. Inferred state probabilities use all the information available until observation t to compute for observation t. Smoothed probabilities use all the information available, called T to compute observation t. Using the estimates of the regime-variable one can then determine when regimes are likely to change and let the model switch.

The inference probabilities $\boldsymbol{\xi}_{t|t}$ and forecast probabilities $\boldsymbol{\xi}_{t+1|t}$ are computed according the following recursions:

$$\boldsymbol{\xi}_{t|t} = \frac{1}{\boldsymbol{\xi}'_{t|t-1}\boldsymbol{d}_t}\boldsymbol{\xi}'_{t|t-1} \odot \boldsymbol{d}_t \text{ and} \tag{15}$$

$$\boldsymbol{\xi}_{t+1|t} = \boldsymbol{P}\boldsymbol{\xi}_{t|t}, \tag{16}$$

where $\boldsymbol{d}_t$ contains the densities of y_t conditional on the regimes and $\boldsymbol{P}$ is the probability matrix as defined in section 3.2.

The smoothed probabilities, $\boldsymbol{\xi}_{t|T}$, are computed according to the following the recursion:

$$\boldsymbol{\xi}_{t|T} = \boldsymbol{\xi}_{t|t} \odot \left(\boldsymbol{P'}(\boldsymbol{\xi}_{t+1|T} \div \boldsymbol{\xi}_{t+1|t})\right) \tag{17}$$

As can be seen in equations 15, 16 and 17, the recursions calculating the inference probabilities and forecast probabilities run forward, meaning they need a starting value, often equal to $\boldsymbol{\xi}_{1|0} = (\zeta, \zeta, ..., \zeta)$, where $\zeta = \frac{1}{\text{amount of regimes}}$. The recursion calculating the smoothed probability however runs backwards starting from $\boldsymbol{\xi}_{T|T}$, which is the final inference probability.

7.1.2 Maximization step

Following the expectation step is the maximization step. In this step the inference, forecast and smoothed probabilities as defined in section 7.1.1 are used to compute estimates of the parameters. The parameter estimation is done by performing maximum likelihood. The parameters that have to be estimated are first of all the coefficients corresponding $\hat{\phi}_j$ to all the predictors $\boldsymbol{X}$ and their variance σ. Then the transition probabilities p_{ij} have to be calculated and finally the new starting value ζ has to be calculated. For the derivations of the optimal values of these parameters, refer to Kole (2019) and section 8.2 of Franses et al. (1998). Where Kole (2019) gives a more detailed derivation of the parameters and Franses et al. (1998) gives the derivation of all coefficients, including those corresponding to all the predictors.

The optimal values $\hat{\phi}_j^{(k)}$ for all regimes $j = 1, ..., J$ in iteration k are obtained as follows:

$$\hat{\phi}_j^{(k)} = \left(\sum_{t=1}^{T} x_t(j) x_t(j)'\right)^{-1} \left(\sum_{t=1}^{T} x_t(j) y_t(j)\right) \text{ where} \tag{18}$$

$$x_t(j) = x_t \xi_{t|T,j}^{(k-1)} \text{ and } y_t(j) = y_t \xi_{t|T,j}^{(k-1)} \tag{19}$$

Next, the optimal values for the variances $\sigma_j^{(k)}$ for all regimes $j = 1, ..., J$ in iteration k are obtained as follows:

$$\sigma_j^{(k)} = \sqrt{\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}(y_t - \hat{\phi}_j^{(k)\prime} x_t)^2 \xi_{t|T,j}^{(k-1)}\right)} \tag{20}$$

The optimal values for the transition probabilities $p_{ij}^{(k)}$ for all regimes $i, j = 1, ...J$ in iteration k are obtained as follows

$$p_{ij}^{(k)} = \frac{\sum_{t=2}^{T} \tilde{p}_{ij,t}}{\sum_{t=2}^{T} \xi_{t-1|T,j}^{(k-1)}} \text{ where} \tag{21}$$

$$\tilde{p}_{ij,t+1} = \xi_{t|t,j}^{(k-1)} \cdot \frac{\xi_{t+1|T,i}^{(k-1)}}{\xi_{t+1|t,i}^{(k-1)}} p_{ij}^{(k-1)} \tag{22}$$

Finally, the starting probabilities $\zeta^{(k)}$ for iteration k are obtained as follows:

$$\zeta^{(k)} = \xi_{t|T,0}^{(k-1)} \tag{23}$$

This algorithm does require a parameter vector with starting values for all parameters. Kole (2019) advises to set the probabilities of remaining in the same regime to high values, for example 0.9, as regimes tend to be persistent. On top of that, Kole (2019) advises to set the distribution parameters of the different regimes to substantially different values.

7.2 Computation of step 2 coefficients

Guérin et al. (2020) explains that there are two ways to compute the coefficients needed for the regressions in step 2 of the MS-3PRF algorithm. They called the coefficients following from these two possible ways, $\hat{\phi}_{A,it}$ and $\hat{\phi}_{B,it}$. Both of these coefficients are very similar in computation, but quite the same. $\hat{\phi}_{A,it}$ and $\hat{\phi}_{B,it}$ are computed as follows:

$$\hat{\phi}_{A,it} = \sum_{j=1}^{M} \boldsymbol{\phi_i}(S_{x_i,t} = j) P(S_{x_i,t} = j|\Omega_t) \tag{24}$$

$$\hat{\phi}_{B,it} = \sum_{j=1}^{M} \boldsymbol{\phi_i}(S_{x_i,t} = j) I(P(S_{x_i,t} = j|\Omega_t)) \tag{25}$$

Here $P(S_{x_i,t} = j|\Omega_t)$ is the smoothed probability, following from the Kim filter, described by Kim (1994), of being in regime j given full information Ω_t. $I(\cdot)$ is an indicator function that selects the regime with the highest smoothed probability. Both estimates are a different version of the MS-3PRF. Guérin et al. (2020) showed that MSS-3PRF seemed to lead to slightly better forecasts.

References

Aastveit, K. A., Carriero, A., Clark, T. E., and Marcellino, M. (2017). Have standard vars remained stable since the crisis? *Journal of applied econometrics*, 32(5):931–951.

Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*, 18(3):338–357.

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications*, volume 418. John Wiley & Sons.

Beaudry, P. and Koop, G. (1993). Do recessions permanently change output? *Journal of Monetary economics*, 31(2):149–163.

Bollen, N. P., Gray, S. F., and Whaley, R. E. (2000). Regime switching in foreign exchange rates:: Evidence from currency option prices. *Journal of Econometrics*, 94(1-2):239–276.

Buffington, J. and Elliott, R. J. (2002). American options with regime switching. *International Journal of Theoretical and Applied Finance*, 5(05):497–514.

Burns, A. F. and Mitchell, W. C. (1947). Measuring business cycles.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Dennison, T. (2018). International investing and the importance of breaking the country box. In *Invest Outside the Box*, pages 229–277. Springer.

Engel, C., Mark, N. C., and West, K. D. (2015). Factor model forecasts of exchange rates. *Econometric Reviews*, 34(1-2):32–55.

Fong, W. M. and See, K. H. (2002). A markov switching model of the conditional volatility of crude oil futures prices. *Energy Economics*, 24(1):71–95.

Franses, P. H., Franses, P. H. B., et al. (1998). *Time series models for business and economic forecasting*. Cambridge university press.

Frömmel, M., MacDonald, R., and Menkhoff, L. (2005). Markov switching regimes in a monetary exchange rate model. *Economic Modelling*, 22(3):485–502.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, pages 763–788.

Goldfeld, S. and Quandt, R. (1973). The estimation of structural shifts by switching regressions. In *Annals of Economic and Social Measurement, Volume 2, number 4*, pages 475–485. NBER.

Greenaway-McGrevy, R., Mark, N. C., Sul, D., and Wu, J.-L. (2018). Identifying exchange rate common factors. *International Economic Review*, 59(4):2193–2218.

Guérin, P., Leiva-Leon, D., and Marcellino, M. (2020). Markov-switching three-pass regression filter. *Journal of Business & Economic Statistics*, 38(2):285–302.

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*, pages 357–384.

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

Harding, D. and Pagan, A. (2002). Dissecting the cycle: a methodological investigation. *Journal of monetary economics*, 49(2):365–381.

Harding, D. and Pagan, A. (2006). Synchronization of cycles. *Journal of Econometrics*, 132(1):59–79.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

Huber, P. J. et al. (1973). Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics*, 1(5):799–821.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, 186(2):294–316.

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

Klaassen, F. (2002). Improving garch volatility forecasts with regime-switching garch. In *Advances in Markov-switching models*, pages 223–254. Springer.

Kole, E. (2019). Markov switching models: An example for a stock market index. *Available at SSRN 3398954*.

Koop, G. and Potter, S. M. (2001). Are apparent findings of nonlinearity due to structural instability in economic time series? *The Econometrics Journal*, 4(1):37–55.

Liu, Z., Waggoner, D. F., and Zha, T. (2011). Sources of macroeconomic fluctuations: A regime-switching dsge approach. *Quantitative Economics*, 2(2):251–301.

MATLAB (2010). *version 9.7.0.1296695 (R2019b) Update 4*. The MathWorks Inc., Natick, Massachusetts.

McKinnon, R. and Schnabl, G. (2003). Synchronised business cycles in east asia and fluctuations in the yen/dollar exchange rate. *World Economy*, 26(8):1067–1088.

Meese, R. A. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of sample? *Journal of international economics*, 14(1-2):3–24.

Mussa, M., Masson, P., Swoboda, A., Jadresic, E., Mauro, P., and Berg, A. (2000). *Exchange rate regimes in an increasingly integrated world economy*, volume 193. Citeseer.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of econometrics*, 131(1-2):445–473.

Psaradakis, Z. and Spagnolo, N. (2003). On the determination of the number of regimes in markov-switching autoregressive models. *Journal of time series analysis*, 24(2):237–252.

Rossi, B. (2013). Exchange rate predictability. *Journal of economic literature*, 51(4):1063–1119.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Terasvirta, T. and Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of applied econometrics*, 7(S1):S119–S136.

Wilfling, B. (2009). Volatility regime-switching in european exchange rates prior to monetary unification. *Journal of International Money and Finance*, 28(2):240–270.