Trinity College

HARTFORD CONNECTICUT

Department of Economics
Trinity College
Hartford, CT
06106 USA

<http://www.trincoll.edu/depts/econ/>

TRINITY COLLEGE DEPARTMENT OF ECONOMICS WORKING PAPER 16-02

Markov Switching in Exchange Rate Models: Will More Regimes Help?

Josh Stillwagon and Peter Sullivan

December 2016

This paper examines the performance of Markov switching models (MSM) of the exchange rate using a data driven approach to determine the number of regimes. The analysis is conducted for the British pound/USD over the last thirty years with alternative specifications from the literature. A noteworthy finding is that there is a close correspondence among the number of regimes that minimizes mean square forecast errors, the number of regimes selected by traditional information criteria (but not Markov switching specific information criteria), and the fewest regimes with well-behaved residuals. Although allowing for more regimes yields improvement over single or two regime models, the MSM is still unable to outperform a random walk.

J.E.L. Codes: F31, C24

Keywords: Exchange rates; Markov switching; Monetary models; Segmented trends

MARKOV SWITCHING IN EXCHANGE RATE MODELS: WILL MORE REGIMES HELP?

Josh Stillwagon^{a,c}

and

Peter Sullivan^{b,c}

November 2016

This paper examines the performance of Markov switching models (MSM) of the exchange rate using a data driven approach to determine the number of regimes. The analysis is conducted for the British pound/USD over the last thirty years with alternative specifications from the literature. A noteworthy finding is that there is a close correspondence among the number of regimes that minimizes mean square forecast errors, the number of regimes selected by traditional information criteria (but not Markov switching specific information criteria), and the fewest regimes with well-behaved residuals. Although allowing for more regimes yields improvement over single or two regime models, the MSM is still unable to outperform a random walk.

Keywords: Exchange rates, Markov Switching, Monetary Models, Segmented Trends

J.E.L. Codes: F31, C24

Word Count: 7494

^aDepartment of Economics, Trinity College, Hartford, CT – joshua.stillwagon@trincoll.edu

^bDepartment of Economics, University of Puget Sound – psullivan@pugetsound.edu

^cInstitute for New Economic Thinking (INET) Program on Imperfect Knowledge Economics (IKE)

1 Introduction

The field of international macroeconomics continues to be plagued by an inability to connect exchange rate movements to movements in fundamentals. While various studies have found improved performance during certain time periods, there has been little consistency (Cheung, Chinn and Pascual 2005). Some hope for monetary models of exchange rate determination has been revived by their longer run performance (Engel, Mark and West 2008), but the consensus has largely been maintained that fundamental factors struggle to account for short term exchange rate movements and even much of the longer term movements. This presents a dramatic source of uncertainty for both policymakers interested in the exchange rate ramifications of their actions and for private sector actors attempting to navigate in an increasingly globalized market economy and financial system.

An important and burgeoning strand of the literature has focused on remedying this shortcoming of exchange rate models by allowing the effects of fundamentals to vary over time. The earliest studies allowed for smoothly evolving parameters (Wolff 1987). Indeed, even Meese and Rogoff's seminal 1983 study allowed for a degree of ad hoc structural change through use of a rolling window for the estimation, but to no avail. Other studies have allowed for non-linear threshold effects where mean reversion back to the fundamental value of the exchange rate only occurs once the divergence is sufficiently large, attributable it is argued to either transaction costs or risk aversion (Kilian and Taylor 2003). Other studies however have offered evidence of much more abrupt changes in the relationship between fundamentals and the exchange rate (Goldberg and Frydman 1996, 2001 and Beckman, Belke, and Kühl 2011). Survey data evidence from foreign exchange dealers similarly provides compelling evidence that the relevant set of explanatory variables changes dramatically over time (Cheung and Chinn 2001).

One method for diagnosing and estimating regime shifts is to employ a Markov switch-

ing (MS) model. This technique found one of its earliest applications in the foreign exchange market, when Hamilton and Engel (1990) strongly rejected a random walk for the exchange rate in favor of a time-varying trend alternating between a fixed positive and negative value. Despite this initial evidence, applications of the Markov switching model with fundamentals have been relatively sparse or confined to a small number of variables. Marsh (2000) for example includes information on the interest rate differential, though he concludes this adds little to the segmented trends model of Hamilton and Engel. Two more recent studies expand the set of variables more extensively. Frömmel, MacDonald and Menkhoff (2005) and De Grauwe and Vansteenkiste (2007) include collectively money supplies, inflation, output, and short and longer term interest rates. Both measure the variables monthly in terms of year over year changes and find evidence of Markov switching between two regimes in the industrial countries.

This study adds to evidence of non-linear regime switching for a monthly sample of the British pound per US dollar spanning over the last thirty years, but differs methodologically from the previous studies in several respects. The first and most important contribution is in testing and allowing for greater than two regimes. Other studies have focused on how the frequency of the data impacts findings of Markov switching, but still maintain the assumption of two regimes (Cheung and Erlandsson 2005). In the segmented trends model and Markov switching models incorporating fundamentals, the conclusion is quite clear that the best approximation in this framework is not merely two regimes over the last three decades, but rather many more, at times approaching or exceeding 10.

The exchange rate disconnect puzzle, originally posited by Meese and Rogoff (1983), is framed not just using in-sample estimation results but most strongly with the poor out-of-sample results of the economic models tested.¹ The results of Frömmel, MacDonald, and

¹For overviews of this literature, see Frankel and Rose 1995, Cheung, Chinn, and Pascual 2005, and Frydman and Goldberg 2007.

Menkhoff (2005) and De Grauwe and Vansteenkiste (2007) both provide evidence that a two regime Markov switching model out performs linear versions of their models in out-of-sample fit, but neither are able to outperform the random walk benchmark.

There are two natural questions for this research. The first is whether allowing for more regimes will improve the out of sample performance. The second is whether the best performing out of sample models are consistent with the optimum deduced by the in-sample criteria. Using minimum mean square error (MSE) for the one-month ahead forecasts as our metric of out-of-sample fit, we find that as more regimes are added the predictive power increases, but only up to a point, just as with measures of in sample fit. More strikingly, the preferred number of regimes based on in-sample information criteria and out-of-sample forecasting performance are quite similar. These results on the improved performance, both in and out of sample, when allowing for more regimes demonstrates that there is clearly much more structural change occurring than simply the two recurrent regimes often assumed.

However, even when we search ex post for the number of regimes which minimizes MSE (creating a very dramatic bias in favor of the MSM), the improvement in predictive power is still insufficient in any of the MSMs to outperform a random walk. Further, inspection of the recursively calculated MSE shows sharp degradations following the 2008 financial crisis. These results indicate that the MS framework may be too constricting or mechanical to sufficiently capture structural change.²

²Frydman and Goldberg's (2007, 2011) theory of Imperfect Knowledge Economics (IKE) strongly argues that this is the case not just with Markov-switching models, but all models that ignore the presence of unforeseeable change.

2 Monetary Models and Markov Switching

Frömmel, MacDonald, and Menkhoff (2005) draw from the real interest differential (RID) model of Frankel (1979) which encompasses both the flexible and sticky price models of exchange rate determination.

$$e_t = c + (m_t - m_t^*) + \beta_1(y_t - y_t^*) + \beta_2(i_t - i_t^*) + \beta_3(b_t - b_t^*) + \varepsilon_t \quad (1)$$

The variable e_t represents the exchange rate specified in terms of the domestic price of foreign currency, m_t is the money supply, y_t is output, i_t is the short term interest rate, and b_t is the long term bond rate, where the superscript * indicates foreign values. Hereafter, it will be understood that all of the explanatory variables are in relative terms. c is a constant, and ε_t is the i.i.d. mean zero error term with constant variance.

For the vast majority of studies, the coefficients are assumed constant in the regressions. Goldberg and Frydman (1996, 2001) however find evidence of discrete breaks in the relationship between fundamentals and the exchange rate using a CUSUM test for structural change.

The Markov switching model (MSM) was popularized by Hamilton (1989). Typically the estimation has allowed for only two regimes, in that case we have:

$$e_t = c_i + \beta_i X_t + \varepsilon_t \quad \text{if } s = s_i \quad (2)$$

$$e_t = c_j + \beta_j X_t + \varepsilon_t \quad \text{if } s = s_j \quad (3)$$

so we have one relationship when in state i and another when in state j . The regime switching is assumed to follow an ergodic and irreducible first order Markov process where the probability of a switch is determined by the current state. This implies that the influence of

all of the past observations for the variables and states is fully encapsulated in the current observation of the state variable. The probability of switching is captured in the matrix of the transition probabilities Π :

$$\Pi = \begin{vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{vmatrix} \quad (4)$$

where p_{ij} is the probability of being in state i next period conditional on being in state j this period. This framework is easily expanded to allow for n -regimes. The Markov switching model provides a rather flexible method ex post to approximate structural change in the data generating process. It allows the data to determine the timing of the shifts, the recurrence of regimes, and in which regimes any given variable may matter and to what degree. There can be asymmetries in the persistence of different regimes, and the expectation of the dependent variable can even be independent of the current state if $p_{11} = (1 - p_{22})$.

3 Segmented Trends Model

Engel and Hamilton (1990) and Engel (1994) strongly rejected a random walk for the exchange rate in favor of their broken trends model where the exchange rate alternated between a fixed upward and downward trend according to a Markov Chain. Their samples however principally included the long dollar depreciation, appreciation, and depreciation that followed the advent of a floating exchange rate. Perhaps this behavior was merely a by-product of markets coming to grips with this new monetary regime. It is of interest then to extend the sample in order to determine whether this statistical approximation is still apt. Further, the modeling may be improved by allowing for a greater number of regimes.

We begin by replicating the initial findings, though now with a longer series, but still with only two states. The equation being estimated is given by:

$$\Delta e_t = \mu_{t,i} + \varepsilon_t \quad (5)$$

where Δe_t is the year over year percentage changes in the exchange rate, $\mu_{t,i}$ is a constant and $i = 1, 2$.³ This implies that the change in the exchange rate varies between two constant values according to a first order Markov process. The results of this estimation are presented in tables 1 and 2.

Even with the extended sample, we obtain similar results to Engel and Hamilton (1990). The exchange rate appears to alter between a fixed positive and negative trend, both of which are statistically significant at the 1% level. The probabilities of a switch are low, below 10% in both cases, so the long swings are quite persistent. The likelihood ratio test of a linear model is rejected with a p-value of 0.0000. Clearly then, the exchange rate appears to behave not as a random walk, but as a more persistent trending process over the sample.

Even this model however appears to be quite inadequate, as the model fit is rather poor and the residuals display a strong amount of serial correlation. Merely two regimes is seemingly omitting some of the variation in the time-varying trend. The question then becomes: How many regimes should we include? Selection based only by model fit would tend to lead to a proliferation of parameters and over-fitting. Instead we focus on the Schwarz (1978) or Bayesian (SC) information criteria, the Akaike (1973) information criteria (AIC) and the Markov switching criteria (MSC) proposed by Smith, Naik, and Tsai (2006). We also observe the residual diagnostics to ensure the absence of misspecification and reliable inference.

As can be seen in table 3, and is well known, the AIC penalizes additional parameters less harshly than the SC, indicating 14 and 13 regimes respectively. Yet both of these criteria rely on regularity conditions that break down when some of the parameters of the model

³The exchange rate data is the end of month value collected from the St. Louis FRED database. The year over year percentage change is calculated as $\Delta e_t = \frac{e_t - e_{(t-12)}}{e_{(t-12)}}$.

are unidentified under the null (Hamilton 2016). The Markov-switching criteria (MSC) proposed by Smith, Naik and Tsai (2006) attempts to deal with this problem, proposing a penalty that is based on the sum of the switching probabilities of each state.

An important note is that when considering high state systems, the MSC test statistic behaves poorly. Higher regimes with lower probabilities, particularly relative to the number of parameters, can yield the result that the denominator of the penalty term approaches zero and eventually reverses sign. This leads to selection of arbitrarily large models. To circumvent this, we only consider MSC values for which the penalty term is increasing with additional parameters. Of the first nine regimes meeting this requirement of an increasing penalty, the local minimum occurs at six regimes. A key finding here is that, according to any of the three metrics, the preferred model contains far greater than two regimes, with the minimum deduced being six.⁴

The residual analysis, reported in table 3, further indicates that at least eleven regimes are required to have a well specified model. Hendry and Krolzig (2001) propose that a p-value threshold of 1% is appropriate for each specification test (for no serial correlation (AR), no ARCH, and Normality) since using 5% would lead to a type I error of rejecting an adequate model based on one of the three tests greater than 5% of the time. Comparing with the results from the information criteria, it appears that the SC most closely matches the residual analysis results for the minimum number of regimes required, while the AIC selects a larger model as expected and the MSC selects a model with poorly behaved residuals.

As a representative illustration, results for a model with fourteen regimes, as indicated by the AIC, have a quite straightforward and symmetric interpretation to each regime as, respectively, small, moderate, large and very large appreciation or depreciation trends. Ac-

⁴This larger range of time-varying trend values is similar to the I(2) cointegration specification of Johansen (1997). Johansen et al. (2010) demonstrate that the Markov switching model is actually a special case of the more general I(2) model's time-varying trend. Stillwagon (2016) also finds evidence of I(2) trends in the exchange rate.

cording to the transition matrix, a very large depreciation trend has a 27.5% chance of transitioning to a large depreciation trend, but no chance of switching immediately to an appreciation.⁵ Similarly for a large depreciation, it has a 10.9% chance of transitioning to a small depreciation trend, but no chance of switching directly to an appreciation. Meanwhile, a very large appreciation has a 16.2% chance of becoming a large appreciation and a 1.7% of becoming a moderate appreciation, but with no chance of directly switching to a depreciation. This "inertia" is likely in part a by-product of working with year over year changes which smooths out the series, though this is customary in the MS literature, but can also be seen as consistent with a twilight period prior to the reversal (Soros 2008).

Extending the analysis to consider the out-of-sample properties of this model, we find that the optimal out of sample model is generally consistent with the optimal in sample model based on the SC. The out of sample estimation procedure estimates the model up to a point in time and then uses these coefficients combined with the actual values for the next period regressors to generate a "prediction" of the change in exchange rate next period. Then the data sample is updated by one period and the procedure is repeated. A more detailed description of this procedure is available in the appendix.

This analysis finds that, just as with the in-sample analysis presented earlier, the fit of the model improves as more regimes are added, as measured by the AIC, SC, and mean square error (MSE). These results are displayed in table 4. The information criteria reported are the average of each recursive estimation of the model for AIC and SC. The correlation between the recursively calculated MSE and the recursive average AIC and SC is .98 and .99, respectively, as reported in table 5. The pattern of the improvement in the models is most evident when viewed graphically, as shown in figure 1. The figure clearly shows that the marginal improvement in fit of the models declines as more regimes are added and

⁵The coefficients and transition probabilities are estimated using Oxmetrics's regime switching class of models for PcGive.

eventually reaches a minimum point at which adding more regimes is harmful. Importantly, this pattern of results is not dependent on the particular end date of the sample. The top panel of figure 13 depicts the recursively calculated mean square error for all fourteen regimes. This figure shows that ranking of the models in terms of mean square error is quite consistent throughout the sample.⁶

What is clear from this analysis, both in-sample and out-of-sample, is that the segmented trends model for this data sample exhibits greater than simply two trends. This conclusion is validated not only by the evidently better fit, which one would expect given additional parameters, but also the information criteria, residual analysis, and perhaps most importantly the out-of-sample performance. The question now is whether this finding can be extended to Markov switching models with economic fundamentals, which are of more direct relevance to the exchange rate disconnect puzzle.

4 Markov Switching Models with Fundamentals

We will now explore the impact of adding fundamental variables to the model, in what we refer to as a Markov-switching model (MSM) with fundamentals. The key question we want to address is not so much whether these model outperform the segmented trends model, although that is still of interest, but whether a similar pattern will hold where the model's performance improves as regimes are added, up to some point, based on in-sample and out-of-sample evaluation.

We use a model including those variables incorporated into the previous analyses by Frömmel, MacDonald and Menkhoff (2005), hereafter FMM. Similar overall results however are obtained with the specification used by De Grauwe and Vansteenkiste (2007). The presented results estimate:

⁶Analysis of the recursively calculated information criteria yield the same result that the ranking of the models is not dependent on the end date of the sample.

$$\Delta e_t = c + \beta_1 \Delta y_t + \beta_2 \Delta m_t + \beta_3 \Delta i_t + \beta_4 \Delta b_t + \varepsilon_t \quad (6)$$

where all right hand side variables are in year over year percentage changes and relative terms so $\Delta x_t = \frac{x_t - x_{(t-12)}}{x_{(t-12)}} - \frac{x_t^* - x_{(t-12)}^*}{x_{(t-12)}^*}$.

Here the short term rate is the 3-month T-bill rate, the long term rate is the 10 year government bond, money supply is M2, and output is measured as industrial production to obtain monthly estimates; all collected from the St. Louis Fed's FRED database. First, let's estimate the linear model for comparison's sake, the results of which are shown in table 6. Looking at the coefficients of the model, only the output variable is statistically significant and the other three variables all take on the wrong sign, though imprecisely estimated. Equally important perhaps, the model fit is quite poor and the residuals experience a high degree of serial correlation, as reported in table 8.

Next we will allow for two regimes, as in Frömmel, MacDonald, and Menkhoff (2005), though with our more recent sample. The results of the two regime MSM with fundamentals are reported in table 7. The coefficient estimates from Frömmel, MacDonald, and Menkhoff (2005) are provided for comparison in the column labeled "FMM Coefficient" in table 7. As can be seen, allowing for a degree of structural change through the two regime MSM greatly improves upon the linear specification, as exemplified by the information criteria reported in table 8. The inclusion of fundamentals also improves upon the two regime segmented trends model as well when comparing these with table 3. For example, an AIC of -2.586 is found in the two regime MSM with fundamentals compared with -2.274 for the two regime segmented trends model.

While clearly an improvement on the linear monetary model and the segmented trends model, and hence a step in the right direction, the model is still quite inadequate, as the residuals, reported in table 8, are quite poorly behaved. This undermines the reliability of

our estimates, and indicates that the fit still leaves something to be desired. We will again see if the situation can be improved through the allowance for a greater number of regimes.

Looking first at the residual analysis of the model, seven regimes are required for the residuals to be well behaved. Turning to the information criteria, we see that a model with seven regimes corresponds to the minimum value of the SC, while the minimum value of the AIC is achieved at ten regimes and the MSC indicates only three regimes, following the same general pattern as we observed for the segmented trends model. As before the residuals of the model selected by the MSC are poorly behaved. Similar to the results from the segmented trends model, allowing for more than two regimes improves the fit of the model in terms of both residual analysis and information criteria. Similar results are obtained when allowing for GARCH effects, which suggests that we are picking up more than merely the commonly found heteroskedasticity over time. This is corroborated by the vastly differing coefficient estimates across regimes, discussed next.

The results of allowing for seven regimes with fundamentals are reported in table 9. It is worth providing some caution though that the interpretability of the Markov results may be suspect given the poor out of sample fit found in the following section. With that said, we will first identify each regime based on its trend in the exchange rate extraneous to the fundamentals. This trend is represented by the constant given the dependent variable is specified as the year over year change in the exchange rate. Regime 0 has an insignificant constant, suggesting there is no exogenous trend. All other regimes have trends significant at 1%, though with varying magnitudes. Regime 1 has a small depreciative trend of the pound (based on a constant of -0.051). This could be viewed as slight pessimism towards the pound. Regime 2 has a similar magnitude appreciative trend of the pound (constant 0.043). Regime 3 has a large depreciative trend outside of the included fundamentals (-0.127), while regime four has a moderate appreciative trend of the pound outside of the fundamentals (0.096). Regime 5 has a very strong depreciative trend (-0.286), while regime

6 has a large appreciative trend of the pound (0.132).

Drawing from the existing literature, there are several possible interpretations for the significant trends outside of the fundamental influences. One possibility is that this is capturing some form of extrapolative or bandwagon behavior possibly related to technical trading (e.g. Frankel and Froot 1990, De Grauwe and Grimaldi 2006). Alternatively, it may be capturing animal spirits or other psychological influences including autonomous expectations about future fundamentals. Lastly, it may reflect persistent effects from unobservable “fundamentals”, non-economic news, microstructure effects, or private information.

Looking at the transition probability matrix of table 12, we can see that the regime with the largest depreciative trend (regime 5) can only transition to the regime with the second largest depreciation (regime 3), whereas the regime with the largest appreciative trend (regime 6) can only transition to the second largest appreciative trend regime (regime 3). This reinforces the interpretation that there is some inertia to expectations deriving from psychology, extrapolative noise-trading, or persistent unobservable fundamentals.

Across the different regimes we observe a large number of (often highly) significant variables. However, the variables which are significant often differ across regimes. This is reminiscent of quotes from the Financial Times highlighted by Bacchetta and van Wincoop (2004):

“The dollar’s resilience in the wake of recent dire US economic data has raised the prospect that the currency market may be experiencing one of its periodic changes in focus” - *Financial Times*, February 11, 2008

“The dollar’s latest stumble . . . came despite optimistic economic data from the US. But analysts said the movement of the US currency was no longer driven by growth fundamentals. All the focus is on the deficit now...” - *Financial Times*, February 11, 2003

Survey evidence from Cheung and Chinn (2001) also supports the notion that traders alter their focus or weight on different variables over time. This has been suggested as a source of the widespread findings of parameter instability in exchange rate models (Wolff 1987, Goldberg and Frydman 1996, Rossi 2006, Sarno and Valente 2009, etc.). This provides some general direction for interpreting our regimes. In addition to the trend external to the incorporated fundamentals, the regimes highlight the fundamentals of focus during that time period.

Regime 0 with no significant trend has a focus on primarily relative long term bond rates, where an increase in the US relative to the UK leads to pound appreciation. This may reflect inflation expectations (as is the interpretation often attached in monetary models) or concerns over debt. Regime 1, with the small dollar appreciative trend, focuses again on relative bond rates, with the same sign and slightly larger magnitude, and real growth differentials, where faster RGDP growth in the US leads to dollar appreciation, as one may expect since it signals rising short term interest rates. Regime 2, with the small pound appreciative trend, focuses on the money supply and short term interest rates, where increases in the US relative to the UK lead to dollar appreciation for both variables. The sign for interest rates is consistent with monetary models and the sign on the relative money supply suggests that increases in the money supply lead to expectations of tightening during this regime, as found by Engel and Frankel (1984). Regime 3, with the large dollar appreciative trend, focuses on real output and money supply growth differentials, with the same sign as previously, but the significant sign on interest rates is reversed, suggesting it is capturing inflation effects. Regime 4, with the moderate pound appreciative trend, focuses on real growth, money and interest rate differentials. The signs on output and money are both negative again, and the interest rate is negative as in the monetary models suggesting it is capturing carry trade effects. Regime 5, with the largest dollar appreciative trend, focuses on money and interest rate effects (both with the negative signs again) but now real growth effects have a positive

sign, seemingly viewed as reflecting trade effects rather than being used to predict monetary policy, perhaps since the monetary effects are well captured in the other two variables. Regime 6, with the large dollar depreciative trend, has both output and money growth with their common signs reflecting expected tightening.

One can begin to grasp why even a seven (or more) regime MS model would be too constricting to capture the degree of structural change present in markets *ex ante*, and therefore why the model, even with this added flexibility, is deficient to the point of failing to provide superior forecasts to a random walk (as seen in the following section). One could imagine that the trends could take on a continuum of values, and any combination of fundamentals may be the focus at any one time. The possible combinations of trends, fundamentals of focus, and their varying weights quickly becomes exceedingly large to capture in a prespecified or mechanical fashion.

4.0.1 Out-of-Sample Analysis

Frömmel, MacDonald and Menkhoff (2005) report that while their two regime Markov model significantly outperforms the linear specification of the model, it is still unable to perform a random walk in out-of-sample predictions. They conclude from this result that "although the MSM captures most of the structural instability in the coefficients, there is still some additional source of time variation left." (Frömmel, MacDonald and Menkhoff 2005, pg. 500) A potential source for the unexplained variation in their model is the presence of additional regimes.

The out of sample results of the recursively estimated MSMs with fundamentals are presented in table 11 and figure 2. Similar to the segmented trends model, adding more regimes lowers the information criteria and mean square error up to a point and, as with the segmented trends model, there is a very high correlation (reported in table 12) between the average AIC, SC and MSE.

The minimum MSE for this model is achieved with seven regimes and we find that the ranking of the models in terms of MSE is largely independent of the particular end date of the sample, as reported in the middle panel of figure ???. Interestingly, the out-of-sample results again seem to correspond quite closely to the in sample results reported earlier - the in-sample SC and residual analysis coincide with the minimum MSE at seven regimes. As before, the ranking of the models in terms of MSE and information criteria is not dependent on the particular end date of the sample, as shown in the second panel of figure ???. The MSE of the model with seven regimes is less than that of the two regime model at statistically significant levels according to the Diebold-Mariano (1995) test.⁷ This suggests some of the unexplained variation found by Frömmel, MacDonald and Menkhoff was in fact due to artificially constraining the number of regimes. However, the mean square error of the FMM model estimated with seven regimes is still significantly greater than that of a random walk. This suggests that misspecification remains in either the economic model or how the empirical model is approximating the structural change in the data generating process.

5 Conclusion

The findings of this paper provide a cautionary note for researchers utilizing Markov switching models of the exchange rate. The first lesson is that in characterizing exchange rates ex post, it is advisable to allow and test for greater than two regimes. In fact, the optimum for both in and out-of-sample fit is attained at seven or more regimes for all models with and without fundamentals. The large number of regimes in the segmented trends model suggests the empirical relevance of the I(2) CVAR for exchange rates, as found by Johansen et al. 2010 and Stillwagon 2016.

A noteworthy result from our analysis is that there is a close correspondence among

⁷The MSE of the random walk is 0.000719 as compared to 0.004273 and 0.001452 for the two regime and seven regime FMM model, respectively.

the number of regimes that minimizes the traditional information criteria, the number of regimes that minimizes MSE, and the most parsimonious model that generates well behaved residuals. The number of regimes selected by the Markov switching specific criteria of Smith, Naik and Tsai (2006), however, selects fewer regimes than that which performs best out of sample and leaves misspecification based on the residual diagnostics.

Lastly, while allowing for increased regimes can improve forecasting, even when searching ex post for the number of regimes that minimizes MSE (creating a dramatic bias in favor of the MSM), it is still insufficient to outperform a random walk. While the MS framework provides a flexible way to approximate structural change ex post, (allowing the data to determine the number and timing of regimes, and in which regimes a given variable may matter and to what degree), it becomes rigid in forecasting (assuming all relevant regimes have been discovered and imposing an exact probability of transition between them) The forecast failure of the MSMs with any number of regimes suggests the importance of unforeseeable structural change; seemingly corroborating Frydman and Goldberg's (2007, 2011) emphasis on genuine Knightian (as opposed to probabilistic) uncertainty. It will be of interest in future research to see whether these lessons hold for other domains where the Markov switching framework has been applied, for example with stock prices, interest rates, and business cycles.

Appendix: Out-of-Sample Testing Procedure

This study evaluates the out-of-sample fit of the model following the tradition of Meese and Rogoff (1983). This method compares the predictive accuracy of the structural economic model to that of a simple random walk, using mean-square error. In the tradition of Meese and Rogoff, this is a prediction exercise rather than forecasting because the actual values of the future X 's (regressors) are used to generate the out-of-sample predictions as opposed to requiring these to be forecasted as well.

Predictions are made by estimating the model up to time t , which generates initial coefficient estimates for the model. These estimates are combined with the actual values of the X 's at time $t+k$, where k is the forecast horizon. Predictions are generated for the one-month horizon. Then t is moved forward by one period and the model is re-estimated and new predictions are generated.

The random walk predictions are generated very simply. It assumes that the best prediction of the exchange rate for any point in the future is given by today's exchange rate. In terms of this out-of-sample exercise, this implies that the random walk prediction made at time t for $t + k$ is given by s_t , for $k = 1$.

From these predictions, forecast error statistics are calculated for both the economic model and the random walk model. In this paper, we evaluate the predictive ability of the model using mean square error statistic (MSE). Statistical significance of the difference between the MSE of the economic model and that of the random walk is estimated using the Diebold-Mariano (1995) test statistic.

Figure 1: Out-of-Sample: Segmented Trends

Average AIC and average SC are measured on the left hand axis, depicted by blue (left) and red (right) bars respectively. MSE is measured on the right hand axis, depicted by the black line. Smaller or more negative values denote better fit.

Figure 2: Out-of-Sample: MSM with Fundamentals

Average AIC and average SC are measured on the left hand axis, depicted by blue (left) and red (right) bars respectively. MSE is measured on the right hand axis, depicted by the black line. Smaller or more negative values denote better fit.

Figure 3: Out-of-Sample Recursive MSE

The two panels report the recursively calculated mean square error (MSE) for the segmented trends (top), MSM with fundamentals (bottom), for 14 and 11 regimes respectively.

Table 1: Traditional Two Regime Segmented Trends Model

	coefficient	Std. Error	σ
Regime 0	0.087***	0.009	0.063
Regime 1	-0.064***	0.008	0.071

*** Denotes statistical significance at the 1%, ** at 5% and * at 10% levels.

Table 2: Transition Matrix of the Two Regime Segmented Trends Model

	Regime 0,t	Regime 1,t
Regime 0, t+1	0.941	0.049
Regime 1, t+1	0.059	0.951

Table 3: Residual Analysis for Segmented Trends Models

Regimes	AR	ARCH	Normality	AIC	SC	MSC*
2	0.000	0.000	0.000	-1.713	-1.7350	-268.15
3	0.000	0.000	0.000	-2.274	-2.221	-462.89
4	0.000	0.000	0.000	-2.680	-2.595	-604.00
5	0.000	0.000	0.013	-2.965	-2.847	-693.31
6	0.000	0.000	0.013	-3.228	-3.057	-772.85
7	0.000	0.000	0.108	-3.337	-3.134	-774.74
8	0.030	0.000	0.001	-3.514	-3.225	-723.42
9	0.000	0.000	0.003	-3.541	-3.189	-629.18
10	0.000	0.000	0.000	-3.643	-3.248	
11	0.304	0.277	0.844	-3.703	-3.275	
12	0.475	0.015	0.521	-3.718	-3.237	
13	0.296	0.026	0.338	-3.773	-3.314	
14	0.405	0.108	0.191	-3.799	-3.297	
15	0.166	0.036	0.370	-3.762	-3.217	

The null hypotheses of the residual tests are no auto correlation, no arch effects and normality, respectively. The MSC is reported only up until the point that the penalty term for adding more regimes remains positive.

Table 4: Out-of-Sample: Segmented Trends

Regimes	Avg AIC	Avg SC	MSE
1	-1.5741	-1.5437	0.007429
2	-2.2357	-2.1597	0.004559
3	-2.6355	-2.5140	0.003087
4	-2.8764	-2.7093	0.002333
5	-3.0757	-2.8639	0.001745
6	-3.1895	-2.9261	0.001642
7	-3.3989	-2.9323	0.001501
8	-3.4394	-3.0028	0.001346
9	-3.3989	-2.9872	0.001307
10	-3.3495	-2.9909	0.001302
11	-3.4365	-2.9557	0.001492
12	-3.4410	-2.9300	0.001259
13	-3.4580	-2.9037	0.001256
14	-3.2692	-2.7595	0.001784

Bold values designate the minimum value for each measure of fit. Smaller or more negative values denote better fit.

Table 5: Out-of-Sample: Segmented Trends Correlation Statistics

	Avg AIC	Avg SC	MSE
Avg AIC	1.0000		
Avg SC	0.9763	1.0000	
MSE	0.9807	0.9943	1.0000

This table reports the correlation statistics between out-of-sample average AIC, SC and MSE reported in table 4.

Table 6: Linear MSM with Fundamentals

	Coefficient	Std. Error
c	-0.001	0.006
Δy_t	-0.831***	0.206
Δm_t	-0.199	0.150
Δi_t	0.002	0.006
Δb_t	-0.013	0.044

*** Denotes statistical significance at the 1%, ** at 5% and * at 10% levels.

Table 7: Two Regime MSM with Fundamentals

	Coefficient	Std. Error	FMM Coefficient
c(0)	0.037***	0.005	0.129***
c(1)	-0.082***	0.008	-0.069***
$\Delta y_t(0)$	-0.051	0.217	-0.540***
$\Delta y_t(1)$	-1.295***	0.185	-0.734***
$\Delta m_t(0)$	-1.367***	0.143	-0.872***
$\Delta m_t(1)$	0.117	0.140	0.205
$\Delta i_t(0)$	-0.020*	0.011	-0.009***
$\Delta i_t(1)$	0.019***	0.004	-0.001
$\Delta b_t(0)$	0.005	0.040	0.281***
$\Delta b_t(1)$	0.060	0.051	0.176***
p00	0.955	0.015	0.959
p11	0.933	0.021	0.970

*** Denotes statistical significance at the 1%, ** at 5% and * at 10% levels.

Table 8: Residual Analysis for MSM with Fundamentals

Regimes	AR	ARCH	Normality	AIC	SC	MSC
1	0.000	0.000	0.759	-1.790	-1.726	-288.08
2	0.000	0.000	0.000	-2.586	-2.447	-557.69
3	0.000	0.000	0.000	-2.990	-2.776	-645.32
4	0.000	0.000	0.206	-3.221	-2.911	-624.26
5	0.000	0.000	0.069	-3.440	-3.045	-406.16
6	0.008	0.000	0.155	-3.500	-3.041	
7	0.115	0.098	0.213	-3.610	-3.054	
8	0.466	0.023	0.043	-3.619	-2.967	
9	0.098	0.574	0.230	-3.671	-2.934	
10	0.411	0.017	0.758	-3.802	-2.990	
11	0.429	0.276	0.282	-3.778	-2.859	
12	0.2262	0.245	0.777	-3.808	-2.836	

The null hypotheses of the residual tests are no auto correlation, no arch effects and normality, respectively. The MSC is reported only up until the point that the penalty term for adding more regimes remains positive.

Table 9: MSM with Fundamentals: Estimates with Seven Regime

	Coefficient	Std. Error		Coefficient	Std. Error
c(0)	0.001	0.004	c(4)	0.096***	0.006
y(0)	-0.164	0.208	y(4)	-0.353**	0.145
m(0)	-0.284*	0.147	m(4)	-0.842***	0.133
i(0)	0.003*	0.002	i(4)	-0.021**	0.009
b(0)	0.085***	0.026	b(4)	0.002	0.025
c(1)	-0.051***	0.004	c(5)	-0.286***	0.015
y(1)	-0.229**	0.110	y(5)	1.223***	0.298
m(1)	-0.108	0.114	m(5)	-1.387***	0.197
i(1)	0.012*	0.007	i(5)	-0.057***	0.021
b(1)	0.120***	0.032	b(5)	-0.020	0.066
c(2)	0.043***	0.005	c(6)	0.132***	0.048
y(2)	-0.021	0.111	y(6)	-1.294***	0.235
m(2)	-0.800***	0.083	m(6)	-1.561**	0.794
i(2)	-0.019***	0.003	i(6)	0.157*	0.089
b(2)	0.006	0.029	b(6)	-0.119*	0.062
c(3)	-0.127***	0.005			
y(3)	-0.242**	0.109			
m(3)	-0.365***	0.083			
i(3)	0.039***	0.012			
b(3)	0.023	0.040			

*** Denotes statistical significance at the 1%, ** at 5% and * at 10% levels.

Table 10: Transition Matrix MSM with Fundamentals, Seven Regimes

	0,t	1,t	2,t	3,t	4,t	5,t	6,t
0,t+1	0.755	0.116	0.124	0.000	0.000	0.000	0.000
1,t+1	0.119	0.718	0.017	0.216	0.000	0.000	0.000
2,t+1	0.126	0.029	0.735	0.000	0.252	0.000	0.000
3,t+1	0.000	0.122	0.000	0.716	0.000	0.138	0.000
4,t+1	0.000	0.014	0.110	0.000	0.696	0.000	0.255
5,t+1	0.000	0.000	0.000	0.068	0.000	0.862	0.000
6,t+1	0.000	0.000	0.014	0.000	0.052	0.000	0.745

The values denote the probability of being in state i next period, $t + 1$, shown in the left column, conditional on being in state j this period, t , shown in the top row.

Table 11: Out-of-Sample: MSM with Fundamentals

Regimes	Average AIC	Average SC	MSE
1	-1.5877	-1.4966	0.007613
2	-2.4742	-2.2767	0.004273
3	-2.8538	-2.5500	0.002990
4	-3.1345	-2.7184	0.002122
5	-3.3478	-2.8162	0.001740
6	-3.4182	-2.7763	0.001455
7	-3.3650	-2.6412	0.001452
8	-3.5395	-2.6881	0.001501

Bold values designate the minimum value for each measure of fit. Smaller or more negative values denote better fit.

Table 12: Out-of-Sample: MSM with Fundamentals Correlation Statistics

	Avg AIC	Avg SC	MSE
Avg AIC	1.0000		
Avg SC	0.9618	1.0000	
MSE	0.9925	0.9803	1.0000

This table reports the correlation statistics between out-of-sample average AIC, SC and MSE reported in table 11.

References

- Akaike, H. (1973). "Information theory as an extension of the maximum likelihood principle". In: *Second International Symposium on Information Theory*. Ed. by N. Petrov and F. Csaki. Akademiai Kiado, Budapest, pp. 1–21.
- Bacchetta, Philippe and Eric van Wincoop (2004). "A Scapegoat Model of Exchange-Rate Fluctuations". In: *The American Economic Review* 94.2, pp. 114–118.
- Beckmann, Joscha, Ansgar Belke, and Michael Kühl (2011). "The Dollar-Euro Exchange Rate and Macroeconomic Fundamentals: A Time-Varying Coefficients Approach". In: *Review of World Economics* 147.1, pp. 11–40.
- Cheung, Yin-Wong and Menzie Chinn (2001). "Currency Traders and Exchange Rate Dynamics: A Survey of the US Market". In: *Journal of International Money and Finance* 20, pp. 439–471.
- Cheung, Yin-Wong and Ulf G. Erlandsson (2005). "Exchange Rates and Markov Switching Dynamics". In: *Journal of Business and Economic Statistics* 23.2, pp. 314–320.
- Cheung, Ying-Wong, Menzie D. Chinn, and Antonio Garcia Pascual (2005). "Empirical Exchange Rate Models of the Nineties: Are Any Fit to Survive?" In: *Journal of Money and Finance* 24, pp. 1150–1175.
- De Grauwe, Paul and Marianna Grimaldi (2006). "Exchange rate puzzles: A tale of switching attractors". In: *European Economic Review* 50.1, pp. 1–33.
- De Grauwe, Paul and Isabel Vansteenkiste (2007). *Exchange Rates and Fundamentals: A Non-Linear Relationship?* Tech. rep. 1, pp. 37–54.
- Dielbold, Francis X. and Roberto S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business and Economics Statistics* 13, pp. 253–263.
- Engel, Charles (1994). "Can the Markov Switching Model Forecast Exchange Rates?" In: *Journal of International Economics*, pp. 689–713.
- Engel, Charles and James D. Hamilton (1990). "Long Swings in the Exchange Rate: Are they in the Data and Do Markets Know It?" In: *American Economic Review*, pp. 689–713.
- Engel, Charles, Nelson C. Mark, and Kenneth D. West (2008). "Exchange Rate Models Are Not as Bad as You Think". In: *NBER Macroeconomics Annual 2007*. Vol. 36. 1. University of Chicago Press, pp. 151–165.
- Frankel, Jeffery A. and Kenneth A. Froot (1990). "Chartists, Fundamentalists and Trading in the Foreign Exchange Market". In: *The American Economic Review* 80.2, pp. 181–185.
- Frankel, Jeffery A. (1979). "On the Mark: A Theory of Floating Exchange Rates Based on Real Interest Differentials". In: *The American Economic Review* 69.4, pp. 610–622.
- Frankel, Jeffery A. and Andrew K. Rose (1995). *A Panel Project on Purchasing Power Parity: Mean Reversion Within and Between Countries*. NBER Working Papers 5006. National Bureau of Economic Research.

- Frömmel, Michael, Ronald MacDonald, and Lukas Menkhoff (2005). "Markov Switching Regimes in a Monetary Exchange Rate Model". In: *Economic Modelling* 22, pp. 485–502.
- Frydman, Roman and Michael D. Goldberg (2007). *Imperfect Knowledge Economics: Exchange Rates and Risk*. Princeton, NJ: Princeton University Press.
- Frydman, Roman and Michael D. Goldberg (2011). *Beyond Mechanical Markets*. Princeton, NJ: Princeton University Press.
- Goldberg, Michael D. and Roman Frydman (1996). "Imperfect Knowledge and Behaviour in the Foreign Exchange Market". In: *The Economic Journal* 106.473, pp. 869–893.
- Goldberg, Michael D. and Roman Frydman (2001). "Macroeconomic Fundamentals and the DM/\$ Exchange Rate: Temporal Instability and the Monetary Model". In: *International Journal of Finance and Economics* 6.4, pp. 421–435.
- Hamilton, James D. (1989). "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle". In: *Econometrica* 57, pp. 357–384.
- Hamilton, James D. (2016). "Macroeconomic Regimes and Regime Shifts". In: *Handbook of Macroeconomics*, Vol. 2. Ed. by John Taylor and Harald Uhlig. University of Chicago.
- Johanson, Søren (1997). "Likelihood Analysis of the I(2) Model". In: *Scandinavian Journal of Statistics* 24.4, pp. 433–462.
- Johanson, Søren, Katerina Juselius, Roman Frydman, and Michael D. Goldberg (2010). "Testing Hypotheses in an I(2) Model with Piecewise Linear Trends: An Analysis of the Persistent Long Swings in the Dmk/\$ rate." In: *Journal of Econometrics* 158.1, pp. 117–129.
- Krolzig, Hans-Martin and David F. Hendry (2001). "Computer automation of general-to-specific model selection procedures". In: *Journal of Economic Dynamics and Control* 25.6, pp. 831–866.
- Meese, Richard and Kenneth D. Rogoff (1983). "Empirical Exchange Rate Models of the Seventies: Do they fit out of sample". In: *Journal of International Economics* 14.2, pp. 3–24.
- Rossi, Barbara (2006). "Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability". In: *Macroeconomic Dynamics* 10.3, pp. 554–579.
- Sarno, Lucio and Giorgio Valente (2009). "Exchange Rates and Fundamentals: Footloose or Evolving Relationship". In: *Journal of the European Economic Association* 7.4, pp. 789–830.
- Schwarz, G. (1978). "Estimating the dimension of a model". In: *Annals of Statistics* 6, pp. 461–464.
- Smith, Aaron, Prasad A. Naik, and Chih-Ling Tsai (2006). "Markov-Switching Model Selection using Kullback-Leibler Divergence". In: *Journal of Econometrics* 134, pp. 553–557.
- Soros, George (2008). *The new paradigm for financial markets: the credit crisis of 2008 and what it means*. Basic Books.
- Stillwagon, Joshua (2016). "Are Risk Premia Related to Real Exchange Rate Swings? Evidence from I(2) CVARs with Survey Expectations". In: *Macroeconomic Dynamics* forthcoming.

Wolff, Christian CP (1987). “Time-varying parameters and the out-of-sample forecasting performance of structural exchange rate models”. In: *Journal of Business and Economic Statistics* 5.1, pp. 87–97.