

WHAT CAUSES THE FORECASTING FAILURE OF MARKOV-SWITCHING MODELS ? A MONTE CARLO STUDY

Marie BESSEC*

Othman BOUABDALLAH†

September 8, 2004

Abstract

This paper explores the forecasting abilities of Markov-Switching models. Although MS models generally display a superior in-sample fit relative to linear models, the gain in prediction remains small. We confirm this result using simulated data for a wide range of specifications. In order to explain this poor performance, we use a forecasting error decomposition. We identify four components and derive their analytical expressions in different MS specifications. The relative contribution of each source is assessed through Monte Carlo simulations. We find that the main source of error is due to the misclassification of future regimes.

Keywords: Forecasting, Regime Shifts, Markov Switching.

JEL classifications: C22,C32,C53.

1 Introduction

Since the seminal paper of Hamilton (1989), there is a great deal of interest in modelling time series that are subject to structural changes using Markov-Switching (MS). The cyclical behaviour of many economic variables has been of particular interest.

Several recent studies use MS models to predict economic series (see for example Clements and Krolzig, 1998, Krolzig, 2004). However, the results are disappointing (see Clements et al., 2004, for a review of the literature in this area). Although MS models give a better in-sample fit relative to linear models, they are usually outperformed by linear models in out-of-sample

*marie.bessec@dauphine.fr, EURISCO, University of Paris Dauphine.

†bouabdal@univ-paris1.fr, EUREQua, University of Paris Panthéon-Sorbonne. We would like to thank F. Bec, T.E. Clark, G. Guerrero and P.Y. Hénin for their helpful comments and suggestions. All remaining errors are ours.

forecasting exercises. Dacco and Satchell (1999) present a theoretical explanation for this bad performance in a fairly simple specification. They consider a model with no autoregressive terms and with a switch on the intercept. They show that only a small misclassification of future regimes, due to the failure to forecast the regime indicator, dramatically deteriorates the predictions of this model.

The aim of this paper is to assess the robustness of this result on a wide range of specifications. To this end, we perform a Monte Carlo study. First, the quality of the linear and non-linear predictions are compared. Second, the forecasting error is decomposed as suggested in Krolzig (2004). The analytical expressions of the four different sources of error are derived and their relative contribution is assessed using simulated data.

We focus on specifications with only a shift in the deterministic part where it is possible to derive analytically optimal predictors (Krolzig, 2004). We consider a wide range of specifications for these models. Representations with a switching intercept (and variance) or a switching mean (and variance) are studied using different sets of parameters¹. In particular, we examine the impact of changes in the persistence and error-variance parameters. For all specifications, we show that the failure to predict the future regimes explains the major part of the total prediction error of the MS models.

The remainder of the paper proceeds as follows. Section 2 introduces the four subclasses of the models under study and reports the expression of the optimal predictor in these specifications. Section 3 describes the simulation procedure and compares the performances of linear and non-linear models in forecasting exercises. Section 4 presents the error decomposition and discusses the simulation results that are based on it. Section 5 gives our concluding remarks.

2 Prediction in MS autoregressive models

Krolzig (2004) shows that analytical expressions for the optimal predictors can be derived in MS-VAR models only if the autoregressive parameters are time-invariant. For this reason, we have chosen to focus in the following sections on four important subclasses of MS-VAR models: specifications with switches only on the intercept (MSI), on the intercept and the variance (MSIH), on the mean (MSM) and on the mean and the variance (MSMH). As an illustrative example, we use the special case of univariate specifications with two regimes and one autoregressive term².

¹These specifications are widely used to capture the dynamics of real variables (Hamilton, 1989, Krolzig and Toro, 2002, Clements and Krolzig, 2003) and financial series (Cecchetti *et al.*, 1990, Engel and Hamilton, 1990, Engel, 1994, Garcia and Perron, 1996, Bidarkota, 2001).

²The general case of MS(m)-VAR(p) is presented in detail in the Appendix A.

2.1 The MSI(H) Model

Let y_t be the time series of interest. Suppose that y_t follows a first autoregressive process with a switch on the intercept (MSI). These switches occur between two states and are governed by an unobservable variable S_t which follows a first-order Markov process and takes the value 1 or 2.

$$y_t = \nu_{s_t} + \alpha y_{t-1} + u_t \quad u_t \sim NID(0, \sigma) \quad (1)$$

Following Krolzig (2004), we can define an unobservable 2×1 state vector ξ_t consisting of two binary indicator variables as $\xi_t = [I(s_t = 1), I(s_t = 2)]'$ and F the transition matrix of the Markov process:

$$F = \begin{pmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{pmatrix}$$

The dynamics of the centered state vector of being in state one, $\zeta_t = \xi_{1t} - \xi_1$, is given by:

$$\zeta_{t+1} = (p_{11} + p_{22} - 1)\zeta_t + v_{t+1} \quad (2)$$

where ξ_1 is the first component of the 2×1 vector of ergodic probabilities $\xi = [P(s_t = 1), P(s_t = 2)]'$ and v_t is a martingale difference sequence.

The state space representation of this MSI(2)-AR(1) process can thus be defined by:

$$\begin{cases} y_t - \mu_y = (\nu_1 - \nu_2)\zeta_t + \alpha(y_{t-1} - \mu_y) + u_t \\ \zeta_{t+1} = \rho\zeta_t + v_{t+1} \end{cases} \quad (3)$$

with $\rho = p_{11} + p_{22} - 1$ and $\mu_y = (1 - \alpha)^{-1}(\nu_1, \nu_2) \xi$.

It follows that the optimal predictor $\hat{y}_{t+h|t}$ is given by:

$$\hat{y}_{t+h|t} - \mu_y = \alpha^h(y_t - \mu_y) + (\nu_1 - \nu_2) \left(\sum_{i=1}^h \alpha^{h-i} \rho^i \right) \hat{\zeta}_{t|t} \quad (4)$$

The second term in (4) represents the contribution of the non-linear part. The weight of this term increases with the shift on the intercept $|\nu_1 - \nu_2|$, the persistence parameters α and ρ , and diminishes with the horizon of prediction h . In the absence of change in the intercept ($\nu_1 = \nu_2$), this equation reduces to the linear optimal predictor $\alpha^h(y_t - \mu_y)$.

Note that this analytical expression also applies for a MSIH(2)-AR(1) process where the variance of u_t depends on the state $u_t/s_t \sim NID(0, \sigma_{s_t})$.

2.2 The MSM(H) Model

Let us now consider an AR(1) process with a switching mean as motivated by Hamilton (1989). The dynamics of a MSM(2)-AR(1) model is described by the following equation:

$$y_t = \mu_{s_t} + \alpha(y_{t-1} - \mu_{s_{t-1}}) + u_t \quad u_t \sim NID(0, \sigma) \quad (5)$$

Using the same notations, the state space representation of this model is given by:

$$\begin{cases} y_t - \mu_y = (\mu_1 - \mu_2)\zeta_t + z_t \\ z_{t+1} = \alpha z_t + u_{t+1} \\ \zeta_{t+1} = \rho \zeta_t + v_{t+1} \end{cases} \quad (6)$$

with z_t the autoregressive component of the process $z_t = y_t - \mu_{s_t}$ and $\mu_y = (\mu_1, \mu_2) \xi$.

It is easy to show in this representation that the optimal predictor $\hat{y}_{t+h|t}$ is obtained as follows:

$$\hat{y}_{t+h|t} - \mu_y = \alpha^h (y_t - \mu_y) + (\mu_1 - \mu_2) (\rho^h - \alpha^h) \hat{\zeta}_{t|t} \quad (7)$$

As above, the MSM predictor consists of two parts: the linear optimal predictor and a second part which takes into account the shifts in the mean. The weight of the last one depends on the magnitude of the shift $|\mu_1 - \mu_2|$ and on the persistence of the regimes ρ relative to the persistence of the process α .

Again, this expression is still valid when we allow for a dependence of the variance on the realized regime s_t (MSMH(2)-AR(1) model).

3 Forecasting failure of MS models

Many studies show the poor performance of non-linear models against the linear counterpart for prediction. We explore the robustness of this result for a wide range of DGPs (MSI, MSIH, MSM and MSMH) and different sets of parameters.

To assess the relative performance of the two competing alternatives for forecasting purposes, we perform Monte Carlo simulations. First, data from one of the four MS processes are generated. Then, the linear and non-linear alternatives are estimated³. The lag order of the linear autoregressive model is selected using the BIC criterion with a maximum lag length of 3. Finally, the predictions are computed into the two models at different horizons $h = 1, \dots, 8$. The predictions are made in an out-of sample context with a rolling forecast origin and the estimated parameters are recalibrated at each iteration. This procedure is replicated 2000 times. We consider samples with 200 observations⁴ and the forecast origin T_f rolls from 160 to $200 - h$ for each horizon h . This exercise is repeated for different values of the transition probability $p_{22} \in \{0.70; 0.85\}$ and of the variance parameter $\sigma \in \{0.3; 0.5\}$. The other parameters are chosen close to the estimates of the Hamilton model of the US GNP growth rate (1989): $\mu_1 = \nu_1 = 1$; $\mu_2 = \nu_2 = -1$; $\alpha = 0.2$; $p_{11} = 0.95$.

³We make use of Warne's code available on <http://texlips.hypermart.net/warne/code.html> to estimate the MSI, MSM and MSIH models.

⁴We remove the first 100 observations of the 300 observations initially generated, in order to avoid the possible effects of the initial conditions.

Table 1: Comparison of models with MAE

	MSI				MSIH		MSM				MSMH	
σ	0.3		0.5		0.3, 0.5		0.3		0.5		0.3, 0.5	
p_{22}	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85
1	0.92	0.88	0.94	0.89	0.92	0.88	0.95	0.91	0.96	0.91	0.95	0.91
2	0.96	0.93	0.97	0.92	0.96	0.92	0.97	0.93	0.98	0.92	0.97	0.93
3	0.98	0.96	0.99	0.94	0.98	0.95	0.99	0.95	0.99	0.95	0.99	0.95
4	1	0.98	1	0.96	1	0.97	1	0.96	1	0.96	1	0.96
5	1	0.99	1	0.98	1	0.98	1	0.98	1	0.97	1	0.98
6	1	0.99	1	0.99	1	0.99	1	0.98	1	0.98	1	0.98
7	1	1	1	0.99	1	0.99	1	0.99	1	0.99	1	0.99
8	1	1	1	1	1	1	1	0.99	1	0.99	1	0.99

Table 2: Comparison of models with RMSE

	MSI				MSIH		MSM				MSMH	
σ	0.3		0.5		0.3, 0.5		0.3		0.5		0.3, 0.5	
p_{22}	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85
1	0.96	0.94	0.96	0.93	0.96	0.93	0.98	0.95	0.98	0.95	0.98	0.95
2	0.98	0.96	0.98	0.95	0.98	0.96	0.99	0.96	0.99	0.96	0.99	0.97
3	0.99	0.98	0.99	0.97	0.99	0.98	0.99	0.98	0.99	0.97	0.99	0.98
4	1	0.99	1	0.98	1	0.99	1	0.98	1	0.98	1	0.98
5	1	0.99	1	0.99	1	0.99	1	0.99	1	0.99	1	0.99
6	1	1	1	0.99	1	1	1	0.99	1	0.99	1	0.99
7	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1

The results are summarized in Tables 1 and 2. We report the relative Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the MS predictor to the linear one⁵. A result inferior to one indicates that the Markov Switching model performs better than the linear alternative and *vice versa*. Several findings emerge from these tables. First, the gain of the non-linear alternative relative to the linear one is rather small, although the data are generated from a MS model. Indeed, the gain never exceeds 12% and shrinks to zero for large horizons (as shown above). Such a result is consistent with findings obtained in previous studies (Clements and Krolzig, 1998, Krolzig, 2004). Second, the comparison of the three DGPs shows that the MSIH displays an enhancement of no more than 12% (with the MAE criteria) at short horizons. At longer horizons, the MSI or MSM specifications provide the best relative performance with a maximum gain of 6% (using the MAE criteria). Third, for each DGP, increasing the variance parameter generally leads to a slight deterioration of the MS prediction. On the contrary, an

⁵We have only reported the results for univariate specifications. However, our findings are still valid in the bivariate case. The corresponding results are given in Appendix C.

increase in the persistence of the regimes improves the relative performance of the non-linear specification up to 6%. This increase also slows down the convergence of the non-linear predictor with the linear one as predicted by equations (4) and (7).

4 Forecasting error decomposition

To explain such a poor performance of the MS specifications, we decompose the forecast error of the non-linear models into four components as suggested by Krolzig (2004).

The prediction error $\hat{e}_{t+h|t} = y_{t+h} - E[y_{t+h} / \Omega_t; \hat{\Theta}]$ associated with the optimal predictor $\hat{y}_{t+h|t}$ can be written as follows:

$$\begin{aligned} \hat{e}_{t+h|t} = & (y_{t+h} - E[y_{t+h} / \underline{s}_{t+h}, \Omega_t; \Theta_0]) \\ & + (E[y_{t+h} / \underline{s}_{t+h}, \Omega_t; \Theta_0] - E[y_{t+h} / \underline{s}_t, \Omega_t; \Theta_0]) \\ & + (E[y_{t+h} / \underline{s}_t, \Omega_t; \Theta_0] - E[y_{t+h} / \Omega_t; \Theta_0]) \\ & + (E[y_{t+h} / \Omega_t; \Theta_0] - E[y_{t+h} / \Omega_t; \hat{\Theta}]) \end{aligned} \quad (8)$$

Θ_0 is the set of actual parameters, $\hat{\Theta}$ the estimated set of parameters and Ω_t the information set available at time t . The first component $\hat{e}_{t+h|t}^{(1)}$ reflects the error we get if we know the exact set of parameters and the dynamics of the Hidden Markov process $\underline{s}_{t+h} = \{s_{t+h}, s_{t+h-1}, \dots, s_{t-1}\}$. This source of uncertainty reduces to the unpredictable Gaussian components $(u_s)_{t < s \leq t+h}$. The second term $\hat{e}_{t+h|t}^{(2)}$ measures the contribution of the regime prediction error, *i.e.* the impact of the misclassification of future values of the Markov process. The third one $\hat{e}_{t+h|t}^{(3)}$ measures the error due to the filter uncertainty, that is the error induced by the filtering process of the past and current states involved in the prediction. These three components are evaluated conditional to the true parameters Θ_0 . The last component $\hat{e}_{t+h|t}^{(4)}$ stands for the parameter uncertainty due to the estimation procedure⁶.

We apply this decomposition in the Monte Carlo design described above. For each DGP analyzed in Section 3, the relative weights of each component in absolute value for the eight horizons are depicted in Figure 1. Several results are worth commenting on. First, the third component $\hat{e}_{t+h|t}^{(3)}$ is found to be insignificant in all specifications and at all horizons. Second, the weight of the estimation error $\hat{e}_{t+h|t}^{(4)}$ remains stable and small over all specifications (10-15%). Hence, the two major sources of forecasting error are due to the Gaussian terms and the misclassification of future states. The relative part of these two terms varies across horizons. The first component is the most important at the first horizon ($h = 1$). For larger h , the second component $\hat{e}_{t+h|t}^{(2)}$ dominates with a weight increasing with the horizon and ranging from 40% to 65%. Such a contribution is positively related to the persistence of the regime. On the contrary,

⁶See the Appendix B for the derivation of each component in the MSI(m)-VAR(p) and MSM(m)-VAR(p) specifications.

it tends to decrease with the volatility. This last result is intuitive: a larger variance gives a heavier weight to the unpredictable component, $\hat{e}_{t+h|t}^{(1)}$.

5 Conclusion

In this paper, we have examined the performances of Markov-Switching models in predicting economic variables that are subject to regime switching.

A simulations-based study has shown that the improvement in the forecast performance is rather small compared to the linear specification and occurs only at short horizons. Checking the relevance of this result for different parameter settings has shown the robustness of this finding. Indeed, changing the persistence parameters and the variability of the process does not significantly affect the forecasting performance of the MS models relative to the linear one.

To explain this result, we have performed a forecasting error decomposition exercise. Four different sources of error have been identified and their relative contribution has been assessed using simulated data. It turns out that the misclassification of future-state realizations explains the failure of MS models in prediction exercises with an average contribution of 60% of the total error.

This result suggests that the prediction enhancements made in the MS models require improving the prediction of the states. This will be the subject of future research.

References

- Bidarkota P.V. (2001), "Alternative Regime Switching Models for Forecasting Inflation", *Journal of Forecasting*, 20(1), 21-35.
- Cecchetti G.C., Lam P-S., Nelson C.M. (1990), "Mean Reversion in Equilibrium Asset Prices", *American Economic Review*, 80(3), 398-418.
- Clements M.P., Franses P.H., Swanson N.R. (2004), "Forecasting Economic and Financial Time-Series with Non-linear models", *International Journal of Forecasting*, 20, 169-183.
- Clements M.P., Krolzig H.-M. (1998), "A Comparison of the Forecast Performance of Markov-Switching and Threshold Autoregressive Models of US GNP", *Econometrics Journal*, 1, C47-C75.
- Clements M.P., Krolzig H.-M. (2003), "Business Cycle Asymmetries: Characterization and Testing Based on Markov-Switching Autoregressions", *Journal of Business and Economic Statistics*, 21(1), 196-211.

Dacco R., Satchell C. (1999), "Why do Regime-Switching Forecast So Badly?", *Journal of Forecasting*, 18, 1-16.

Engel C. (1994), "Can the Markov Switching Model Forecast Exchange Rates?", *Journal of International Economics*, 36(1-2), 151-165.

Engel C., Hamilton J.D. (1990), "Long Swings in the Dollar: Are They in the Data and Do Markets Know it?", *American Economic Review*, 80(4), 689-713.

Garcia R., Perron P. (1996), "An Analysis of the Real Interest Rate Under Regime Shifts", *Review of Economics and Statistics*, 78(1), 111-125.

Hamilton J.D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 57, 357-384.

Krolzig H.-M. (2004), "Predicting Markov-Switching Vector Autoregressive Processes", *Journal of Forecasting* (In press).

Krolzig H.-M., Toro J. (2002), "Classical and Modern Business Cycle Measurement: the European Case", *Working Paper, Fundacion Centro de Estudios Andaluces, No. 2002/20*.

APPENDIX

A Optimal predictors

A.1 MSI-VAR model

If the variance and autoregressive parameters of a MS-VAR model are regime-invariant $A_{j,s_t} = A_j$ for $j \in \{1, \dots, p\}$, there exists a linear state space representation. For a MSIH(m)-VAR(p) model, this representation can be written as follows:

$$\begin{cases} y_t - \mu_y = \mathcal{M}\zeta_t + A_1(y_{t-1} - \mu_y) + \dots + A_p(y_{t-p} - \mu_y) + u_t \\ \zeta_{t+1} = \mathcal{F}\zeta_t + v_{t+1} \end{cases}$$

where $\mu_y = (I_K - A_1 - \dots - A_p)^{-1}(\nu_1, \dots, \nu_m)$ ξ is the unconditional mean of y_t , $\mathcal{M} = (\nu_1 - \nu_m, \dots, \nu_{m-1} - \nu_m)$ and $\mathcal{F} = \begin{pmatrix} p_{1,1} - p_{m,1} & \dots & p_{m-1,1} - p_{m,1} \\ \vdots & & \vdots \\ p_{1,m-1} - p_{m,m-1} & \dots & p_{m-1,m-1} - p_{m,m-1} \end{pmatrix}$ is a $(m-1) \times (m-1)$ matrix.

Let us consider the VAR(1) representation of the VAR(p) process. Denoting \underline{x}_t the $Kp \times 1$ vector defined as $\underline{x}_t = (x_t \ x_{t-1} \ \dots \ x_{t-p+1})'$ where x_t is a $K \times 1$ vector, the state space representation can be rewritten as:

$$\begin{cases} \underline{y}_t - \bar{\mu} = \mathcal{H}\zeta_t + A(\underline{y}_{t-1} - \bar{\mu}) + \underline{u}_t \\ \zeta_{t+1} = \mathcal{F}\zeta_t + v_{t+1} \end{cases}$$

where $A = \begin{pmatrix} A_1 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ 0 & & I_K & 0 \end{pmatrix}$ is a $Kp \times Kp$ matrix, $\bar{\mu} = E(\underline{y}_t)$ and $\mathcal{H} = (\mathcal{M} \ 0 \ \dots \ 0)'$ is a $Kp \times (m-1)$ matrix.

It follows that the optimal predictor $\hat{y}_{t+h|t}$ is given by:

$$\hat{y}_{t+h|t} - \mu_y = \left(\sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} \mathcal{F}^i \right) \hat{\zeta}_{t|t} + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu})$$

with $J_{n,np} = (I_n \ 0_n \ \dots \ 0_n)$ a $n \times np$ matrix.

A.2 MSM-VAR model

The state space representation of a MSM(m)-VAR(p) model is given by:

$$\begin{cases} y_t - \mu_y = \mathcal{M}\zeta_t + z_t \\ \underline{z}_{t+1} = A\underline{z}_t + u_{t+1} \\ \zeta_{t+1} = \mathcal{F}\zeta_t + v_{t+1} \end{cases}$$

where $\mu_y = (\mu_1, \dots, \mu_m)$ ξ is the unconditional mean of y_t , $\mathcal{M} = (\mu_1 - \mu_m, \dots, \mu_{m-1} - \mu_m)$ and $z_t = y_t - \mu_y - \mathcal{M}\zeta_t$.

In a MSM(m)-VAR(p) process, the optimal predictor $\hat{y}_{t+h|t}$ is given by:

$$\hat{y}_{t+h|t} - \mu_y = J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}) + \left(\mathcal{M} \mathcal{F}^h J_{(m-1),(m-1)p} - J_{K,Kp} A^h \underline{\mathcal{M}} \right) \hat{\zeta}_{t|t}$$

where $\underline{\mathcal{M}} = I_p \otimes \mathcal{M}$.

B Error Decomposition

B.1 MSI-VAR model

In a MSI(m)-VAR(p) model, the expression of the optimal predictor for the estimated set of parameters is given by:

$$\hat{y}_{t+h|t} = \hat{\mu}_y + \left(\sum_{i=1}^h J_{K,Kp} \hat{A}^{h-i} \hat{\mathcal{H}} \hat{\mathcal{F}}^i \right) \hat{\zeta}_{t|t} + J_{K,Kp} \hat{A}^h (\underline{y}_t - \hat{\bar{\mu}})$$

where $\hat{\theta}$ denotes the estimate of the parameter θ .

The total prediction error is given by:

$$\hat{e}_{t+h|t} = y_{t+h} - E(y_{t+h} | \Omega_t; \hat{\Theta}) = y_{t+h} - \hat{y}_{t+h|t}$$

This error can be decomposed into four components:

$$\hat{e}_{t+h|t} = \hat{e}_{t+h|t}^1 + \hat{e}_{t+h|t}^2 + \hat{e}_{t+h|t}^3 + \hat{e}_{t+h|t}^4$$

- First component (measures the effect of the Gaussian error):

$$\hat{e}_{t+h|t}^1 = y_{t+h} - E(y_{t+h} | s_{t+h}, \dots, s_t, \Omega_t; \Theta_0) = y_{t+h} - \hat{y}_{t+h|t}^1$$

$$\text{with } \hat{y}_{t+h|t}^1 = \mu_y + \sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} \zeta_{t+i} + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}).$$

- Second component (measures the effect of future regime misclassifications):

$$\hat{e}_{t+h|t}^2 = E(y_{t+h} | s_{t+h}, \dots, s_t, \Omega_t; \Theta_0) - E(y_{t+h} | s_t, \Omega_t; \Theta_0) = \hat{y}_{t+h|t}^1 - \hat{y}_{t+h|t}^2$$

$$\text{with } \hat{y}_{t+h|t}^2 = \mu_y + \left(\sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} \mathcal{F}^i \right) \zeta_t + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}).$$

We then deduce:

$$\hat{e}_{t+h|t}^2 = \sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} (\zeta_{t+i} - \mathcal{F}^i \zeta_t)$$

This component is proportional to the error made in predicting the future states $(\zeta_{t+i} - \mathcal{F}^i \zeta_t)$, $i = 1, \dots, h$.

- Third component (due to the error in detecting the current regime):

$$\hat{e}_{t+h|t}^3 = E(y_{t+h} | s_t, \Omega_t; \Theta_0) - E(y_{t+h} | \Omega_t; \Theta_0) = \hat{y}_{t+h|t}^2 - \hat{y}_{t+h|t}^3$$

$$\text{with } \hat{y}_{t+h|t}^3 = \mu_y + \left(\sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} \mathcal{F}^i \right) \hat{\zeta}_{t/t} + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}).$$

It follows that:

$$\hat{e}_{t+h|t}^3 = \sum_{i=1}^h J_{K,Kp} A^{h-i} \mathcal{H} \mathcal{F}^i (\zeta_t - \hat{\zeta}_{t/t})$$

$\hat{e}_{t+h|t}^3$ is related to the filtering error $(\zeta_t - \hat{\zeta}_{t/t})$.

- Fourth component (error due to the estimation process):

$$\hat{e}_{t+h|t}^4 = E(y_{t+h} | \Omega_t; \Theta_0) - E(y_{t+h} | \Omega_t; \hat{\Theta}) = \hat{y}_{t+h|t}^3 - \hat{y}_{t+h|t}$$

B.2 MSM-VAR model

Now, the optimal predictor $\hat{y}_{t+h|t}$ is given by:

$$\hat{y}_{t+h|t} = \hat{\mu}_y + \left(\hat{\mathcal{M}} \hat{\mathcal{F}}^h J_{(m-1),(m-1)p} - J_{K,Kp} \hat{A}^h \hat{\mathcal{M}} \right) \hat{\zeta}_{t/t} + J_{K,Kp} \hat{A}^h (\underline{y}_t - \hat{\mu})$$

In the same way, we can decompose the forecast error into four components:

- First component (the Gaussian error):

$$\hat{e}_{t+h|t}^1 = y_{t+h} - E(y_{t+h} | \underline{s}_{t+h}, \dots, \underline{s}_t, \Omega_t; \Theta_0) = y_{t+h} - \hat{y}_{t+h|t}^1$$

$$\text{with } \hat{y}_{t+h|t}^1 = \mu_y + \mathcal{M} J_{(m-1),(m-1)p} \underline{\zeta}_{t+h} - J_{K,Kp} A^h \underline{\mathcal{M}} \underline{\zeta}_t + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}) \text{ and } \underline{s}_t = \begin{pmatrix} s_t & s_{t-1} & \dots & s_{t-p+1} \end{pmatrix}'.$$

- Second component (misclassification of future regimes):

$$\hat{e}_{t+h|t}^2 = E(y_{t+h} | \underline{s}_{t+h}, \dots, \underline{s}_t, \Omega_t; \Theta_0) - E(y_{t+h} | \underline{s}_t, \Omega_t; \Theta_0) = \hat{y}_{t+h|t}^1 - \hat{y}_{t+h|t}^2$$

$$\text{with } \hat{y}_{t+h|t}^2 = \mu_y + (\mathcal{M} \mathcal{F}^h J_{(m-1),(m-1)p} - J_{K,Kp} A^h \underline{\mathcal{M}}) \underline{\zeta}_t + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}).$$

It follows that:

$$\hat{e}_{t+h|t}^2 = \mathcal{M} (\zeta_{t+h} - \mathcal{F}^h \zeta_t)$$

- Third component (the filtering error):

$$\hat{e}_{t+h|t}^3 = E(y_{t+h} | \underline{s}_t, \Omega_t; \Theta_0) - E(y_{t+h} | \Omega_t; \Theta_0) = \hat{y}_{t+h|t}^2 - \hat{y}_{t+h|t}^3$$

$$\text{with } \hat{y}_{t+h|t}^3 = \mu_y + (\mathcal{M} \mathcal{F}^h J_{(m-1),(m-1)p} - J_{K,Kp} A^h \underline{\mathcal{M}}) \hat{\zeta}_{t/t} + J_{K,Kp} A^h (\underline{y}_t - \bar{\mu}).$$

We then deduce:

$$\hat{e}_{t+h|t}^3 = (\mathcal{M}\mathcal{F}^h J_{(m-1),(m-1)p} - J_{K,Kp} A^h \underline{\mathcal{M}}) (\underline{\zeta}_t - \hat{\underline{\zeta}}_{t/t})$$

Note that this error is now dependent on the filtering of the current as well as of the $p-1$ past regimes.

- Fourth component (due to the estimation error):

$$\hat{e}_{t+h|t}^4 = E(y_{t+h} | \Omega_t; \Theta_0) - E(y_{t+h} | \Omega_t; \hat{\Theta}) = \hat{y}_{t+h|t}^3 - \hat{y}_{t+h|t}$$

C Results for bivariate MS-processes

This appendix assesses the robustness of our results for bivariate MS-processes. To this aim, we simulate data from MS-VAR models with a switch on the intercept or on the mean and eventually on the variance parameters⁷. The transition probabilities are chosen as in the univariate design: $p_{11} = 0.95$ and $p_{22} = \{0.70, 0.85\}$. The other parameters of the equation $i = \{1, 2\}$ are given by: $\mu_1^i = \nu_1^i = 1$; $\mu_2^i = \nu_2^i = -1$ and $A = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}$. At last, the errors of the two equations are supposed to be uncorrelated and of equal variance.

We report in the following the results obtained in the first equation of the VAR. The results for the second one - very similar - are not given here.

The comparison exercise shows the same qualitative results as those obtained in the univariate specifications. The predictive gain of the MS specification relative to the linear one is small and non significant for horizons larger than one. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the MS predictor relative to the linear one obtained in the first equation are reported in Tables 3 and 4. A result inferior to one still indicates that the Markov-Switching model performs better than the linear alternative and *vice versa*. We note that the gain in prediction is a slightly smaller than in the univariate case and converges more quickly to zero.

The explanation for this failure is the same as in the linear framework. Figure 2 gives the decomposition of the error prediction for the first variable of the VAR. Again the two major sources of forecasting error are due to the Gaussian terms and the misclassification of future states. The first component is the most important at the first horizon ($h = 1$). For larger h , the second component dominates with a weight increasing with the horizon and ranging from 40% to 65%.

⁷Given the computational burden, we do not consider the bivariate MSMH specifications.

Table 3: Comparison of models with MAE

	MSI				MSIH		MSM			
σ	0.3		0.5		0.3, 0.5		0.3		0.5	
p_{22}	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85
1	0.94	0.91	0.95	0.92	0.94	0.91	0.97	0.95	0.98	0.95
2	0.97	0.96	0.97	0.94	0.97	0.94	0.99	0.96	0.99	0.95
3	0.99	0.99	0.99	0.96	0.99	0.98	0.99	0.97	1	0.96
4	1	1	0.99	0.98	0.99	1	1	0.98	1	0.97
5	1	1.01	1	0.99	1	1.01	1	0.99	1	0.98
6	1	1.01	1	1	1	1.01	1	0.99	1	0.99
7	1	1.01	1	1	1	1.01	1	0.99	1	0.99
8	1	1.01	1	1	1	1.01	1	0.99	1	0.99

Table 4: Comparison of models with RMSE

	MSI				MSIH		MSM			
σ	0.3		0.5		0.3, 0.5		0.3		0.5	
p_{22}	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85	0.70	0.85
1	0.97	0.95	0.97	0.95	0.97	0.95	0.99	0.98	0.99	0.97
2	0.98	0.97	0.98	0.96	0.98	0.97	0.99	0.98	0.99	0.97
3	0.99	0.98	0.99	0.98	0.99	0.98	1	0.99	1	0.98
4	0.99	0.99	0.99	0.99	0.99	0.99	1	0.99	1	0.99
5	0.99	0.99	1	0.99	0.99	0.99	1	1	1	0.99
6	0.99	0.99	1	1	0.99	0.99	1	1	1	0.99
7	1	0.99	1	1	1	0.99	1	1	1	1
8	1	0.99	1	1	1	0.99	1	1	1	1

Figure 1: Error decomposition results

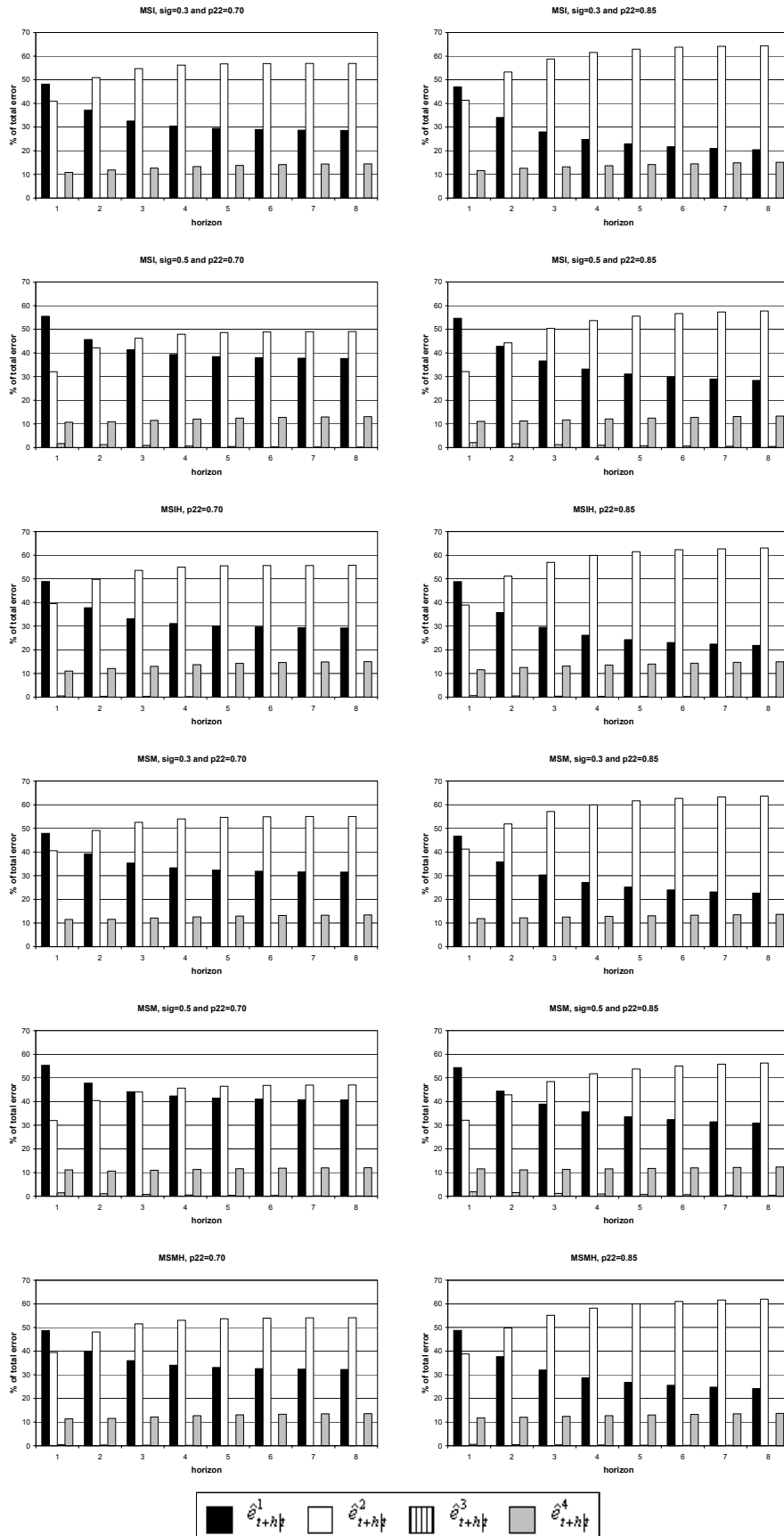


Figure 2: Error decomposition results in the bivariate case (first equation)

