Predicting Markov-Switching Vector Autoregressive Processes

Hans-Martin Krolzig *
Department of Economics
and Nuffield College, Oxford.

April 19, 2000

Abstract

While there has been a great deal of interest in the modelling of non-linearities and regime shifts in economic time series, there is no clear consensus regarding the forecasting abilities of these models. In this paper we develop a general approach to predict multiple time series subject to Markovian shifts in the regime. The feasibility of the proposed forecasting techniques in empirical research is demonstrated and their forecast accuracy is evaluated.

Keywords: Forecasting, Regime Shifts, Structural Breaks, Causality, Predictability, Intercept Correction, Business Cycle, Markov Switching.

JEL classification: C53, E37, C22, C32.

1 Introduction

In recent years the importance of structural breaks in explaining forecasting failures has become evident. Consequently, for an econometric theory of forecasting to deliver relevant conclusions about empirical forecasting, it must be based on assumptions that adequately capture this aspects of the real world. A theory of economic forecasting applicable to time series subject to deterministic breaks has been presented in Clements and Hendry (1999). For breaks that recur in a systematic, stochastic pattern, modelling the regime-switching nature of economic processes might result in forecast devices superior to time-invariant linear models and traditional robustifying methods like differencing, intercept correction and multistep estimation.

While there has been a great deal of interest in modelling non-linear features of economic time series, there appears to have been little attempt to investigate statistical forecasting using regime-switching models, and it is that lacuna this paper seeks to fill. One of the puzzles associated with the forecasting performance of non-linear times series models in general, and regime-switching models in particular is that, when compared to linear models, a superior in-sample fit does not result in superior

*I benefited greatly from discussions with Mike Clements, Rob Engle, David Hendry, Søren Johansen, Massimiliano Marcellino, Grayham Mizon, Hashem Pesaran and Juan Toro. Useful comments were received from conference participants at the ESRC Workshop on Forecasting, Oxford, the ESEM99, Santiago de Compostela, the American Wintermeeting of the Econometric Society, Boston, and seminar audiences at the EUI, Florence, the University of Osnabrück and the University of Manchester. The recent structure of the paper was initiated while I visited the EUI, Florence. I am very grateful to the EUI for its hospitality, and to the UK Economic and Social Research Council under grant L116251015 for its financial support. I am also grateful to Jim Hamilton for graciously providing me with his data. All the computations reported in this section were carried out with the MSVAR class in Ox: see Doornik (1998) and Krolzig (1998a).

forecasts (see, *inter alia*, Clements and Krolzig, 1998, and Dacco and Satchell, 1999). Another feature found in empirical investigations and simulation studies is that the relative forecast performance of regime-switching models depends on the regime present at the time the forecast is made (see, *inter alia*, Clements and Smith, 1999, and Pesaran and Potter, 1997) .These issues will be addressed in this paper and theoretical explanations will be offered.

We propose an econometric theory of predicting economic time series when the data generating mechanism incorporates endogenous structural change by being linear conditional on a particular regime, but subject to shifts in the deterministic factors. Thus the focus is on the predictability of Markov-switching vector autoregressive (MS-VAR) processes as the property of a stochastic process in relation to an information set. We derive the optimal predictor, we show that its properties depend on (i) the significance of regime shifts, (ii) the persistence of the regime generating process, (iii) the asymmetry of the regime generating process and (iv) the interaction with the autoregressive dynamics. The results obtained allow to derive parametric conditions under which the optimal predictor shrinks to a linear prediction rule.

The paper proceeds as follows: In section 2 the Markov-switching (MS) model is introduced as the framework for the following analysis and the general approach to predict MS processes is laid out. We will discuss the central concepts of prediction density, optimal predictor, unpredictability and Granger causality of regimes. The optimal predictor of MS regression models is derived in section 3. In section 4 we discuss the prediction of MS time series processes. The concepts of unpredictability and Granger causality of regimes are found to be crucial for the information value of the statistical regime inference. The illustrative examples in the second part of the paper demonstrate the feasibility of the proposed forecasting facilities. Their forecast accuracy is evaluated and compared to those of linear and non-linear alternatives. In section 5 we consider the Hamilton (1989) model of the US business cycle which allows us to exemplify the derived forecasting techniques and to test the forecast performance of the exact model specification that spearheaded the recent interest in MS-AR models relative to linear and non-linear alternatives. Two methods of analysis are presented: an empirical forecast accuracy comparison of the Hamilton model with linear autoregressive models, and a Monte Carlo study. We then show that the forecasting performance of the MS model can be hugely improved by allowing for a third 'high-growth' regime, and simultaneously modelling US output and employment growth in an MS-VAR.

2 The Framework

2.1 Markov-Switching Models

By allowing for changes in regime of the process generating the time series, the MS-VAR model has been proposed as an alternative to the constant-parameter, linear time-series models of the earlier Box and Jenkins (1970) modelling tradition. The general idea behind this class of regime-switching models is that the parameters of a, say, K-dimensional vector time series process $\{y_t\}$ depend upon an *unobservable* regime variable $s_t \in \{1, \ldots, M\}$, which represents the probability of being in a particular state of the world.

$$p(y_t|Y_{t-1}, X_t, s_t) = \begin{cases} f(y_t|Y_{t-1}, X_t; \theta_1) & \text{if } s_t = 1 \\ \quad\vdots & \\ f(y_t|Y_{t-1}, X_t; \theta_M) & \text{if } s_t = M. \end{cases} \tag{1}$$

where $Y_{t-1} = \{y_{t-j}\}_{j=0}^{\infty}$ denotes the history of y_t and X_t are strongly exogenous variables; θ_m is the parameter vector associated with regime m.

A complete description of the statistical model requires the formulation of a mechanism that governs the evolution of the stochastic and unobservable regimes on which the parameters of (1) depend. Once a law has been specified for the states s_t, the evolution of regimes can be inferred from the data. In MS models the regime-generating process is an ergodic Markov chain with a finite number of states defined by the transition probabilities:[1]

$$p_{ij} = \Pr(s_{t+1} = j | s_t = i), \quad \sum_{j=1}^{M} p_{ij} = 1 \quad \forall i, j \in \{1, \ldots, M\}. \tag{2}$$

More precisely, it is assumed that s_t follows an ergodic M-state Markov process with an irreducible transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & & \vdots \\ p_{M1} & \cdots & p_{MM} \end{bmatrix}.$$

Thus, the probability which regime is in operation at time t conditional on the information at time $t-1$ only depends on the statistical inference on s_{t-1}, $\Pr(s_t | Y_{t-1}, X_t, S_{t-1}) = \Pr(s_t | s_{t-1})$. As the MS model also nest models with once-and-for-all structural breaks, it might be used to detect permanent breaks. In this case the matrix of transition probabilities could look as follows:

$$\mathbf{P} = \begin{bmatrix} 1-p_{12} & p_{12} & 0 & \cdots & 0 & 0 \\ 0 & 1-p_{23} & p_{23} & & 0 & 0 \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ 0 & 0 & 0 & & 1-p_{M-1,M} & p_{M-1,M} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

A major advantage of the MS model is its flexibility in modelling time series subject to regime shifts. Theoretically all parameters of the conditional model can be made dependent on the state s_t of the Markov chain. In this paper we consider different specifications of the conditional processes $p(y_t | Y_{t-1}, X_t, s_t)$. The focus, however, is on stochastic processes exhibiting shifts in the deterministic factors:

The *Markov-switching regression model* is defined as

$$y_t = \begin{cases} X_t \beta_1 + u_{t,} & u_t | s_t \sim \mathsf{NID}(0, \Sigma_1) \quad \text{if } s_t = 1 \\ \quad \vdots & \\ X_t \beta_M + u_{t,} & u_t | s_t \sim \mathsf{NID}(0, \Sigma_M) \quad \text{if } s_t = M \end{cases} \tag{3}$$

where X_t is a $(K \times R)$ regressor matrix of exogenous variables such that v_t and u_t are innovation processes.

The most general form of a *Markov-switching vector autoregressive (MS-VAR) process* is given by

$$y_t = \nu(s_t) + A_1(s_t) y_{t-1} + \ldots + A_p(s_t) y_{t-p} + u_t, \quad u_t | s_t \sim \mathsf{NID}(0, \Sigma(s_t)), \tag{4}$$

[1]There is evidence that in some instances the assumption of fixed transition probabilities p_{ij} should be relaxed, and models with time-varying and duration-dependent transition probabilities have been considered (see, for example, Diebold, Rudebusch and Sichel, 1993, Diebold, Ohanian and Berkowitz, 1994, Filardo, 1994, Lahiri and Wang, 1994, and Durland and McCurdy, 1994). The former are modeled as logistic functions (to bound the probabilities between 0 and 1) of economic variables.

where the presample values $y_0, \ldots, y_{1-p}$ are fixed. The parameter shift functions $\nu(s_t)$, $A_1(s_t), \ldots, A_p(s_t)$, and $\Sigma(s_t)$ describe the dependence of the parameters on the realized regime s_t, for example:

$$\nu(s_t) = \begin{cases} \nu_1 & \text{if } s_t = 1, \\ & \vdots \\ \nu_M & \text{if } s_t = M. \end{cases}$$

As most empirical forecast errors are due to shifts to the deterministic factors (see Hendry, 1999),we will focus on models with shifts in the level of the process. From the class of MS-VAR processes, these model allow an easy analytical access to the properties of the optimal multi-step predictor. This is a great advantage when compared to other regime-switching models, where numerical integration or simulations methods have to be used to calculate (e.g. see Clements and Smith, 1997, for the case of SETAR models).

A VAR with regime shifts in the *mean* is called an MSM(M)-VAR(p) process:

$$y_t - \mu(s_t) = \sum_{k=1}^{p} A_k \left(y_{t-k} - \mu(s_{t-k})\right) + u_t, \quad u_t | s_t \sim \mathsf{NID}(0, \Sigma). \tag{5}$$

If the regime shifts affect the *intercept* of the VAR, this is called an MSI(M)-VAR(p) process:

$$y_t = \nu(s_t) + \sum_{k=1}^{p} A_k \, y_{t-k} + u_t, \quad u_t | s_t \sim \mathsf{NID}(0, \Sigma). \tag{6}$$

The difference between these alternatives that after a shift in regime, the transition to the new (conditional) mean is smooth in an MSI-VAR and once-and-for-all in an MSM-VAR. One special feature of these models is the linearity of their state-space representation. Thus, MSM-VAR and MSI-VAR processes represent the subclass of MS-VAR processes for which optimal predictor can be derived analytically and computationally effective algorithms can be constructed easily.

2.2 The Markov Property of MS-VAR Processes

The prediction of MS-VAR processes uses the is *Markov* property of the joint process $\{(\mathbf{y}_t', s_t)'\}$ of the regime variable s_t and the stacked vector of observed variables $\mathbf{y}_t = (y_t', y_{t-1}', \ldots, y_{t-p+1}')'$, i.e. the relevant information concerning the evolution of the system output in the future $(\mathbf{y}_{t+h}', s_{t+h})', h > 0$ is completely provided by the actual state $(\mathbf{y}_t', s_t')'$, while the past reveals no additional information.

Conditional on the history of regimes, $S_t = \{s_{t-j}\}_{j=0}^{\infty}$, the density of $\mathbf{y}_t$ entails the Markov property,

$$p(\mathbf{y}_t | Y_{t-1}; S_t) = p(\mathbf{y}_t | \mathbf{y}_{t-1}; s_t),$$

since it only depends on the distribution of the error term u_t which is independent of Y_{t-1}. However, the marginal process $\mathbf{y}_t$ generally is not Markovian[2]

$$p(\mathbf{y}_t | Y_{t-1}) = \sum_{m=1}^{M} p\left(\mathbf{y}_t | \mathbf{y}_{t-1}, s_t = m\right) \Pr\left(s_t = m | Y_{t-1}\right) \overset{a.e.}{\neq} p(\mathbf{y}_t | \mathbf{y}_{t-1}).$$

Former observations of $\mathbf{y}_t$ reveal information on the unobservable s_t which then affects the predictive probability density of y_t:

$$\Pr(s_t | Y_{t-1}) \overset{a.e.}{\neq} \Pr(s_t | \mathbf{y}_{t-1}).$$

[2]This suggests the existence of finite-order mixed VARMA representation of these processes (see Krolzig, 1997, chapter 3).

Only if the regimes are not autocorrelated, $\Pr(s_t|s_{t-1}) = \Pr(s_t)$, the Markov property of $\mathbf{y}_t$ would be re-established as $p(\mathbf{y}_t|Y_{t-1}; s_{t-1}) = p(\mathbf{y}_t|Y_{t-1})$. In this case the regime variable s_t is said to be unpredictable.

2.3 The Prediction Density

In contrast to Gaussian models where interval forecasts and forecast regions can be derived on the basis of the conditional mean $\hat{y}_{t+h|t}$ and the h-step MSPE matrix $\mathbf{\Sigma}_{t+h|t} = \mathsf{E}\left[(y_{t+h} - \hat{y}_{t+h|t})(y_{t+h} - \hat{y}_{t+h|t})'|\Omega_t\right]$, the conditional first and second moments are not sufficient to determine the conditional density of y_{t+h} given the information set Ω_t. The prediction density $p(y_{t+h}|\Omega_t)$ is a mixture of normals, $f(y_{t+h}|s_{t+h} = j, \Omega_t)$, with weights given by the predicted regime probabilities $\Pr(s_{t+h} = j|\Omega_t)$:

$$\begin{aligned} p(y_{t+h}|\Omega_t) &= \sum_{j=1}^{M} \Pr(s_{t+h} = j|\Omega_t) p(y_{t+h}|s_{t+h} = j, \Omega_t) \\ &= \sum_{j=1}^{M} \left\{ \sum_{i=1}^{M} \Pr(s_{t+h} = j|s_t = i) \Pr(s_t = i|\Omega_t) \right\} p(y_{t+h}|s_{t+h} = j, \Omega_t) \end{aligned} \tag{7}$$

The predicted probability densities are non-normal and thus in general neither symmetric, homoscedastic, nor regime invariant. Properties which can be barely captured by linear time series models. Thus researchers interested in density forecasts will even stronger benefit from using MS-VAR models (see the discussion in Pesaran and Potter, 1997).

For example the one-step prediction density is given by

$$p(y_{t+1}|\Omega_t) = \sum_{j=1}^{M} \left\{ \sum_{i=1}^{M} p_{ij} \Pr(s_t = i|\Omega_t) \right\} p(y_{t+1}|s_{t+1} = j, \Omega_t) \tag{8}$$

where, in the case of an MS regression model, $p(y_{t+1}|s_{t+1} = m, \Omega_t)$ is Gaussian with expectation $X_{t+1}\beta_m$ and variance Σ_m. Even if the variances are regime-invariant, the prediction density is conditionally heteroscedastic as the weights, $\Pr(s_{t+1} = j|\Omega_t) = \sum_{i=1}^{M} p_{ij} \Pr(s_t = i|\Omega_t)$, of the mixture of normals density are time-varying and predictable.

Although the preceding calculations have been straightforward, in practice it is rather complicated to construct interval forecasts analytically. Therefore the following analysis focuses on optimal point prediction. [3]

2.4 The Optimal Predictor

A forecasting rule is any systematic operational procedure for making statements about future events. For the mean square prediction error (MSPE) criterion,

$$\min_{\hat{y}} \Sigma_{t+h|t} = \min_{\hat{y}} \mathsf{E}\left[(y_{t+h} - \hat{y})(y_{t+h} - \hat{y})' \middle| \Omega_t \right],$$

the optimal predictor $\hat{y}_{t+h|t}$ is given by the conditional mean for a given information set Ω_t :

$$\hat{y}_{t+h|t} = \mathsf{E}[y_{t+h}|\Omega_t]. \tag{9}$$

[3] A forecasting tool which incorporates parameter uncertainty, non-normality of the prediction error, as well as non-linearities of the process, is the Gibbs sampler proposed in Kim and Nelson (1998) and Krolzig (1998b). A main advantage of the Gibbs sampler is the feasibility of generating forecasting intervals by producing the predicted density of y_{t+h} given Y_t.

In contrast to linear models, the MSPE optimal predictor $\hat{y}_{t+h|t}$ usually does not have the property of being a linear predictor if the true data-generating process is non-linear. Unlike many non-linear models, the conditional mean can easily be derived analytically if the autoregressive parameters are regime-invariant.

From the conditional density (8), the MSPE-optimal one-step predictor results as follows:

$$\hat{y}_{t+1|t} = \mathsf{E}[y_{t+1}|\Omega_t] = \sum_{j=1}^{M} \Pr(s_{t+1} = j|\Omega_t)\mathsf{E}\left[y_{t+1}|s_{t+1} = j, \Omega_t\right]. \tag{10}$$

The prediction error $\hat{e}_{t+h|t} = y_{t+h} - \mathsf{E}[y_{t+h}|\Omega_t]$ associated with the optimal predictor $\hat{y}_{t+h|t}$ is given by:

$$\begin{aligned}\hat{e}_{t+h|t} &= (y_{t+h} - \mathsf{E}[y_{t+h}|s_{t+h}, \Omega_t]) + (\mathsf{E}[y_{t+h}|s_{t+h}, \Omega_t] - \mathsf{E}[y_{t+h}|\Omega_t]) \\ &= (y_{t+h} - \mathsf{E}[y_{t+h}|s_{t+h}, \Omega_t]) + (\mathsf{E}[y_{t+h}|s_{t+h}, \Omega_t] - \mathsf{E}[y_{t+h}|s_t, \Omega_t]) + (\mathsf{E}[y_{t+h}|s_t, \Omega_t] - \mathsf{E}[y_{t+h}|\Omega_t])\end{aligned}$$

The h-step prediction error can be decomposed into three components reflecting three causes of uncertainty: (i) the Gaussian innovations $y_{t+h} - \mathsf{E}[y_{t+h}|s_{t+h}, \Omega_t]$ affecting the measurement equation, and the regime prediction errors, which consists of (ii.a) the accumulated, unpredictable errors to the regime generating process and (ii.b) the error in detecting the actual state of the Markov chain due to the signal-extraction problem (filter uncertainty), $\mathsf{E}[y_{t+h}|s_t, \Omega_t] - \mathsf{E}[y_{t+h}|$. If parameters have to be estimated as it is usually the case in practice, another term enters due to parameter uncertainty.

2.5 Predicting the Markov Chain

As we have seen in the preceding discussion, predicting future regime probabilities conditional on the available information set is the major step in implementing optimal predictors for the time series of interest. A useful framework to work with is the VAR(1) representation of the hidden Markov chain. Let ξ_t be an $(M \times 1)$ vector whose ith element is unity when $s_t = i$ and zero otherwise. Then the hidden Markov chain can be written as (see Krolzig, 1997):

$$\xi_{t+1} = \mathbf{F}\xi_t + v_{t+1}, \tag{11}$$

where v_{t+1} is a martingale difference sequence (MDS), $\xi_t = \left[\begin{array}{ccc} I(s_t = 1) & \cdots & I(s_t = M) \end{array}\right]'$ is the unobservable state vector consisting of M binary indicator variables and $\mathbf{F} = \mathbf{P}'$ is called the transition matrix. This representation allows the application of the well-established theory of forecasting linear systems to the problem of calculation future regime probabilities.

By using the linearity of the transition equation (11), the one-step prediction of ξ_{t+1} follows as:

$$\hat{\xi}_{t+1|t} - \bar{\xi} = \mathbf{F}(\hat{\xi}_{t|t} - \bar{\xi}). \tag{12}$$

where $\hat{\xi}_{t|t}$ is the vector of filtered regime probabilities:

$$\Pr(s_t|y_t, X_t) = \frac{p(y_t|s_t, Y_{t-1}, X_t)\Pr(s_t|Y_{t-1}, X_t)}{p(y_t|Y_{t-1}, X_t)}$$

In practice the vector of filtered regime probabilities $\hat{\xi}_{t|t}$ can be calculated recursively with the filtering algorithm of Hamilton (1988, 1989):

$$\hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{\eta_t' \hat{\xi}_{t|t-1}} = \frac{1}{L}\left(\prod_{j=0}^{t-1} \mathrm{diag}(\eta_{t-j})\mathbf{F}\right)\xi_0$$

with $\eta_t = \begin{bmatrix} p(y_t|X_{t,}s_t = 1) = f_u(y_t - X_t\beta_1) \\ \vdots \\ p(y_t|X_{t,}s_t = M) = f_u(y_t - X_t\beta_M) \end{bmatrix}$.

By using the adding-up restriction, $\mathbf{1}\prime_M \xi_t = 1$ for all t, we can reduce the dimension of the Markov chain by one. Thus the state of the Markov chain at time t can be recorded by the $([M-1] \times 1)$ regime vector

$$\zeta_t = \begin{bmatrix} \xi_{1t} - \bar{\xi}_1 \\ \vdots \\ \xi_{M-1,t} - \bar{\xi}_{M-1}. \end{bmatrix}.$$

The stochastic process of the regime vector ζ_t can again be represented as a VAR(1) process:

$$\zeta_t = \mathcal{F}\zeta_{t-1} + v_t, \tag{13}$$

where the $([M-1] \times [M-1])$ transition matrix $\mathcal{F}$ has eigenvalues within the unit circle,

$$\mathcal{F} = \begin{bmatrix} p_{11} - p_{M1} & \cdots & p_{M-1,1} - p_{M1} \\ \vdots & & \vdots \\ p_{1,M-1} - p_{M,M-1} & \cdots & p_{M-1,M-1} - p_{M,M-1} \end{bmatrix},$$

and v_t is a non-Gaussian MDS.

According to (12) the optimal predictor is given by

$$\hat{\zeta}_{t+h|t} = \mathcal{F}\hat{\zeta}_{t|t}. \tag{14}$$

In the case of a two-regime model, the state vector is scalar and $\mathcal{F} = [p_{11} - p_{21}]$ This is illustrated further in the following example.

Example 1. *Consider a two state Markov chain such that the unobservable state vector is* $\xi_t = \begin{bmatrix} I(s_t = 1) & I(s_t = 2) \end{bmatrix}'$. *Define* $\zeta_t = \xi_{1t} - \bar{\xi}_1$ *being* $1 - \bar{\xi}_1$ *if the regime is* 1 *and* $-\bar{\xi}_1$ *otherwise.* $\bar{\xi}_1 = p_{21}/(p_{12} + p_{21})$ *is the unconditional probability of regime* 1*. Invoking the unrestricted VAR(1) representation of a Markov chain*

$$\zeta_t = (p_{11} - p_{21})\zeta_{t-1} + v_t. \tag{15}$$

where v_t *is a martingale difference sequence. Thus (15) is a non-Gaussian autoregressive process characterized by its persistence parameter* $\rho = p_{11} + p_{22} - 1$*. By using the law of iterated predictions, we first derive the forecast of* ζ_{t+h} *conditional on* ζ_t*.*

$$\zeta_{t+h} = \rho^h \zeta_t$$

Then, the expectation operator is again applied to the just derived expressions, but now conditional on the sample information Y_t*. Then predictions of the Markov chain are given by:*

$$\hat{\zeta}_{t+h|t} = \rho^h \hat{\zeta}_{t|t}$$

where $\hat{\zeta}_{t|t} = \mathsf{E}(\zeta_t|Y_t) = \hat{\xi}_{1t|t} - \bar{\xi}_1$ *is the mean-adjusted filtered probability of being in regime* 1 :

$$\hat{\zeta}_{t|t} = \left[1 + \frac{\eta_2}{\eta_1}\left(\left[\hat{\zeta}_{t|t-1} + \bar{\xi}_1\right]^{-1} - 1\right)\right]^{-1} - \bar{\xi}_1.$$

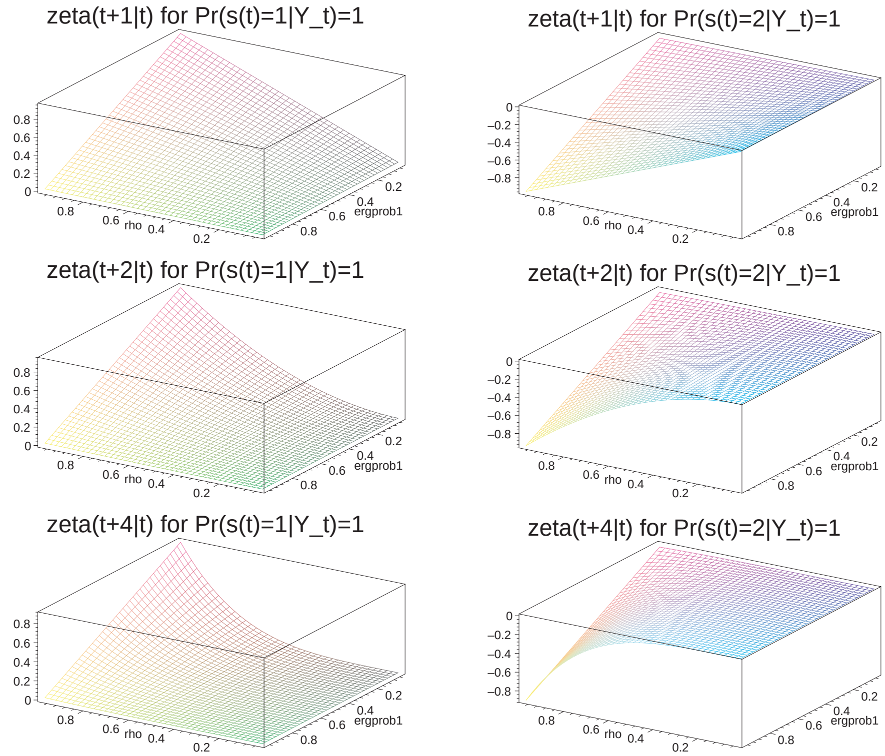

Figure 1 Prediction of the unobserved state variable for $0 < \rho < 1$.

The predictive value of the regime inference, at the time when the prediction is made, depends positively on the regime persistence, i.e. the eigenvalue(s) of $\mathcal{F}$, which is ρ in the two-regime case. The value is the higher, the less likely the reconstructed regime is. This is illustrated in figure 1 for the case $M = 2$ and gives the value of the mean-adjusted predicted probability of regime 1, $\hat{\zeta}_{t+h|t}$, conditional on probabilities of one for regime 1 and 2, respectively, as a function of the ergodic probability of regime 1 and the regime persistence parameter $\rho = p_{11} + p_{22} - 1$.

2.6 Predictability and Granger Causality of Regimes

Based on our previous results, we can formally define the concepts of predictability and Granger causality of regimes. As we will show in the rest of the paper, these concepts are essential for whether the information derived by modelling the regime switching nature of the DGP over the sample period will matter for predictions, or whether we can stick to a linear prediction rule. We will also derive the corresponding parameter restrictions on the MS-VAR.

Predictability is a property of a stochastic process in relation to an information set:

Definition 1. *The regime generating process* $\{s_t\}$ *is said to be* **unpredictable** *iff the regimes are serially independent such that for all* t *and* $s_{t+1}, s_t = 1, \ldots, M$:

$$\Pr(s_{t+1}|s_t) = \Pr(s_{t+1}).$$

In terms of the transition matrix F in (11), unpredictability follows from $F = \bar{\xi}\mathbf{1}'$. Note that the definition is formulated in terms of the unobservable, latent regime variable s_t. However this property of the regime generating process affects immediately also the predicted inference which is based on $\Pr(s_{t+h}|Y_{t-1}) = \Pr(s_{t+h})$.

Proposition 1. *If the regime generating process* $\{s_t\}$ *is unpredictable***,** *detecting regime shifts has no predictive value for future regimes*

$$\Pr(s_{t+1}|s_t) = \Pr(s_{t+1}) \quad \Rightarrow \quad \Pr(s_{t+h}|\Omega_t) = \Pr(s_{t+h}).$$

Proof. From (7) follows that the predicted regime probabilities, $\Pr(s_{t+h}|Y_t)$, are is a linear function of the filtered regime probabilities, $\Pr(s_t|Y_t)$:

$$\Pr(s_{t+h} = j|\Omega_t) = \sum_{i=1}^{M} \Pr(s_{t+h} = j|s_t = i) \Pr(s_t = i|\Omega_t).$$

As unpredictability implies that $\Pr(s_{t+h} = j|s_t = i) = \Pr(s_{t+h} = j)$, we have that all future expectations are given by the ergodic regime probabilities $\Pr(s_{t+h} = j)$:

$$\Pr(s_{t+h} = j|\Omega_t) = \Pr(s_{t+h} = j) \left\{ \sum_{i=1}^{M} \Pr(s_t = i|\Omega_t) \right\} = \Pr(s_{t+h} = j).$$

■

The predictability of a regime generating process $\{s_t\}$ relative to the available information will have important implications for the predictability of the time series vector y_t. In the following the concept of Granger (1969) causality is adapted for the prevailing non-Gaussian framework to measure the link between the latent regime variable s_t and the observed variable y_t.

Definition 2. *The regime* $\{s_t\}$ *is said to be non-causal for the observed times series vector* $\{y_t\}$ *in a* **strict** *sense iff for all* t

$$p(y_{t+1}|\Omega_t; \lambda) = p(y_{t+1}|\Omega_t, \xi_t; \lambda).$$

For well-specified models, causal information always can be shown to be useful, and can produce better forecasts. Forecasting tends to focus on the optimal predictor. In this case it is sufficient to consider weak non-causality which we define as non-causality in the mean:

Definition 3. *The regime* $\{s_t\}$ *is said to be non-causal for the observed times series vector* $\{y_t\}$ *in a* **weak** *sense iff for all* t

$$\mathsf{E}[y_{t+1}|\Omega_t; \lambda] = \mathsf{E}[y_{t+1}|\Omega_t, \xi_t; \lambda].$$

Under conditions of weak non-causality, the optimal predictor shrinks to a simple linear autoregressive forecasting rule which could be estimated by OLS. Thus the forecast accuracy of a linear prediction rule can not be beaten by the optimal predictor of the MS-VAR. However, if the focus is on the predicted densities and not on the optimal predictor itself, then modelling the MS-VAR can be important for forecasting.

In the next section we will show that in the case of a Markov-switching regression models, unpredictability of the regimes implies strong Granger non-causality. Thus a linear forecasting rule is optimal and the predicted densities are mixtures of normals with time-invariant weights given by the ergodic regime probabilities.

3 Prediction of Markov-Switching Regression Models

As the formal framework for the statistical analysis and prediction we work with the state-space representation of the Markov-switching model. The state-space representation consists of the measurement equation (16), which is the regression model, and the transition equation (17), which is the VAR(1) representation of the Markov chain:

$$
\begin{aligned}
y_t &= X_t \mathbf{B} \xi_t + u_t, && (16)\\
\xi_{t+1} &= \mathbf{F} \xi_t + v_{t+1}, && (17)
\end{aligned}
$$

where u_t is Gaussian, v_{t+1} is an MDS. Note that ss-obs) follows straightforward from (3) by setting $\mathbf{B} = [\beta_{1,} \dots, \beta_M]$.

By ignoring the parameter estimation problem, the MSPE–optimal forecast can be generated by the conditional expectation (9) of the of the state-space representation (16)/(17). Consider first the simplest case of an MS-regression model, where the parameter matrix $\mathbf{B}$ is known and the regressor matrix X_t is deterministic. Then the expectation of y_{t+1} conditional on the regime ξ_{t+1} and the observations Y_t is given by:

$$
\mathsf{E}[y_{t+1}|X_{t+1}, \xi_{t+1}] = X_{t+1}\ \mathbf{B}\ \xi_{t+1}, \tag{18}
$$

where we have used the unpredictability of the Gaussian innovation process u_t, *i.e.* $\mathsf{E}[u_{t+1}|X_{t+1}, \xi_{t+1}] = 0$. Thus, in case of anticipation of regime m, the optimal predictor would be $X_{t+1}\beta_m$.

Having forecasts for the predetermined variables, the major task is to forecast the evolution of the hidden Markov chain which can be achieved by using (12). Inserting the predicted state vector $\hat{\xi}_{t+1|t}$ into equation (18) yields the one-step predictor $\hat{y}_{t+1|t}$:

$$
\begin{aligned}
\hat{y}_{t+1|t} &= \mathsf{E}[y_{t+1}|X_{t+1}] = X_{t+1}\ \mathbf{B}\ \hat{\xi}_{t+1|t}\\
&= X_{t+1}\bar{\boldsymbol{\beta}} + X_{t+1}\ \mathbf{B}\ \mathbf{F}(\hat{\xi}_{t|t} - \bar{\xi}). && (19)
\end{aligned}
$$

Starting with the one–step prediction formula (19), general predictions can be derived iteratively as long as the elements of X_{t+h} are uncorrelated with the state vector ξ_{t+h}

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \mathsf{E}[X_{t+h}|Y_t]\ \mathbf{B}\ \hat{\xi}_{t+h|t}\\
&= \mathsf{E}[X_{t+h}|Y_t]\bar{\boldsymbol{\beta}} + \mathsf{E}[X_{t+h}|Y_t]\mathbf{B}\mathbf{F}^j\left(\hat{\xi}_{t|t} - \bar{\xi}\right) && (20)
\end{aligned}
$$

where $\bar{\boldsymbol{\beta}} = \mathsf{E}[\mathbf{B}\ \xi_{t+h}]$ is the unconditional parameter vector $\boldsymbol{\beta}_t = \mathbf{B}\ \xi_{t+h}$.

Since the developed forecasting devices employ the MSPE optimal predictor, the problem of non-Gaussian densities is not involved in the following analysis. However note that due to the non-normality of the innovations of the regime-generating process v_{t+1}, the inferences $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$ depend on the information set Y_t in a non-linear fashion. Hence, in contrast to Gaussian state-space models, the one-step prediction of $y_{t+1|t}$ cannot be interpreted as a linear projection.

The size of the dynamic intercept correction depends on the significance of the regime shift, i.e. the $\beta_m - \beta_M$ collected in the matrix $\mathcal{B}$, the persistence of regimes, i.e. the eigenvalues of $\mathcal{F}$ and the filtered state of the system when the forecast is made:

$$
\hat{y}_{t+h|t} = \mathsf{E}[X_{t+h}]\bar{\boldsymbol{\beta}} + \mathsf{E}[X_{t+h}]\mathcal{B}\mathcal{F}^j\hat{\zeta}_{t|t} \tag{21}
$$

Consider finally the implications of unpredictable regime on predicting MS regression models:

Proposition 2. *In MS regression models, the regime s is Granger non-causal or the observed times series vector y (in a* **strict** *sense) iff the regime is unpredictable:*

$$
\hat{\xi}_{t+h|t} = \bar{\xi} \quad \Rightarrow \quad \hat{y}_{t+h|t} = X_{t+h}\bar{\beta}
$$

The MS regression model can be represented by a time-invariant **linear** *model with heteroscedastic non-Gaussian errors w_t (mixture of normals):*

$$y_t = X_t\bar{\beta} + w_t, \quad f(w_t) = \sum_{m=1}^{M} \bar{\xi}_m f_u \left(w_t - X_t(\beta_m - \bar{\beta})\right)$$

4 Prediction of Markov-Switching Vector Autoregressive Processes

In a time series framework, where lagged endogenous variables are included in the regressor matrix X_{t+j}, equation (20) does not hold in general. This results is due to the correlation of the lagged endogenous variables contained in X_t with the regime vector ξ_t in models with regime-dependent autoregressive dynamics:

$$\mathsf{E}[X_{t+h}\ \mathbf{B}\ \xi_{t+h}|Y_t] \neq \mathsf{E}[X_{t+h}|Y_t]\ \mathbf{B}\ \mathsf{E}[\xi_{t+h}|Y_t].$$

However in MS-VAR with time-invariant autoregressive parameters, which are in the center of our discussion, this problem does not occur. We will therefore consider these separately.

For clarity of exposition consider the MS(M)-VAR(1) model

$$\begin{aligned} y_t &= A(\xi_t)y_{t-1} + u_t, \\ \xi_t &= F\xi_{t-1} + v_t, \end{aligned}$$

where u_t is $\mathsf{NID}(0, \Sigma)$ and v_t is an MDS. It follows that

$$\begin{bmatrix} \xi_{1t}y_t \\ \vdots \\ \xi_{Mt}y_t \end{bmatrix} = \begin{bmatrix} p_{11}A_1 & \cdots & p_{M1}A_1 \\ \vdots & & \vdots \\ p_{1M}A_M & \cdots & p_{MM}A_M \end{bmatrix} \begin{bmatrix} \xi_{1t-1}y_{t-1} \\ \vdots \\ \xi_{Mt-1}y_{t-1} \end{bmatrix} + \varepsilon_t,$$

or $\psi_t = \mathbf{\Pi}\psi_{t-1} + \varepsilon_t$ where $\psi_t = \xi_t \otimes y_t$ and ε_t is an MDS. Hence the multi-step prediction of ψ_t is given by:

$$\mathsf{E}\left[\psi_{t+h}|\psi_t\right] = \mathbf{\Pi}^h\psi_t. \tag{22}$$

As $y_t = \sum_{i=1}^{M} \xi_{it}y_t$ we have that the conditional expectation y_t can be derived based on the representation (22):

$$\begin{aligned} \mathsf{E}\left[y_{t+h}|\Omega_t\right] &= \sum_{i=1}^{M} \mathsf{E}\left[\xi_{it+h}y_{t+h}|\Omega_t\right] = (\mathbf{1}'_M \otimes \mathbf{I}_K)\mathsf{E}\left[\psi_{t+h}|\Omega_t\right] \\ &= (\mathbf{1}'_M \otimes \mathbf{I}_K)\Pi^h \mathsf{E}\left[\psi_t|\Omega_t\right] = (\mathbf{1}'_M \otimes \mathbf{I}_K)\Pi^h \left(\mathsf{E}\left[\xi_t|\Omega_t\right] \otimes y_t\right). \end{aligned}$$

Thus the optimal predictor is given by

$$\hat{y}_{t+h|t} = (\mathbf{1}'_M \otimes \mathbf{I}_K)\Pi^h \left(\hat{\xi}_{t|t} \otimes y_t\right). \tag{23}$$

The properties of the optimal predictor depend on the properties of the matrix $\mathbf{\Pi}$. As forecast errors are mainly due to shifts in the level and the drift of economic time series, we will focus in the following on MS-VAR models with shifts in the mean or intercept.

4.1 MSI–VAR Processes

If the variance parameters, $\Sigma(s_t) = \Sigma$, and the autoregressive parameters are regime-invariant, $A_j(s_t) = A_j$ for $j = 1, \ldots, p$, there exists a *linear* state-space representation of MS-VAR models (see Krolzig, 1997, ch.2) This also implies that MSI(M)-VAR(p) and MSM(M)-VAR(p) processes possess a *linear* time-invariant MA(∞) representation of y_t in its Gaussian and Markov chain innovations .

For the MSI-VAR model defined in (6), the state-space representation is the point to start with. For example, we have in case of an MSI(M)-VAR(1):

$$\begin{aligned} y_t &= \mathbf{M}\xi_t + A y_{t-1} + u_t, \quad \mathbf{M} = [\nu_1, \ldots, \nu_M]\,. && (24)\\ \xi_{t+1} &= \mathbf{F}\xi_t + v_{t+1}, && (25) \end{aligned}$$

such that $X_t = \left[\mathsf{I}_K \otimes (1, y'_{t-1})\right]$ and $\beta_m = [\nu'_M, \alpha']'$ where $\boldsymbol{\alpha} =\mathsf{vec}(\mathbf{A}')$.

MSPE-optimal forecasts can be derived by applying the conditional expectation to the measurement equation,

$$y_t - \mu_y = \mathcal{M}\zeta_t + A_1\left(y_{t-1} - \mu_y\right) + \ldots + A_p\left(y_{t-p} - \mu_y\right) + u_t,$$

where we used that $\mu_y = \left(\mathsf{I}_K - A_1 - \ldots - A_p\right)^{-1}(\nu_1, \ldots, \nu_M)\bar{\xi}$. Again, the lagged endogenous variables $\mathbf{y}_{t-1}$ and the regime vector ζ_t enter additively. Thus, the optimal h-step predictor is given by

$$\hat{y}_{t+h|t} - \mu_y = \mathcal{M}\hat{\zeta}_{t+h|t} + A_1\left(\hat{y}_{t+h-1|t} - \mu_y\right) + \ldots + A_p\left(\hat{y}_{t+h-p|t} - \mu_y\right), \tag{26}$$

with $\mathcal{M} = (\nu_1 - \nu_M, \ldots, \nu_{M-1} - \nu_M)$ and $\hat{\zeta}_{t+h|t}$ follows from (12). To derive a closed form solution for $\hat{y}_{t+h|t}$, we use the stacked VAR(1) representation of a VAR(p) process. Denoting $\mathbf{y}_t = (y'_t, \ldots, y'_{t-p+1})'$, equation (6) can then be rewritten as

$$\mathbf{y}_t - \mu_y = \mathcal{H}\,\zeta_t + J\mathbf{A}\left(\mathbf{y}_{t-1} - \bar{\boldsymbol{\mu}}\right) + u_t, \tag{27}$$

where $\mathbf{A} = \begin{bmatrix} A_1 & \ldots & A_{p-1} & A_p \\ \mathsf{I}_K & & \mathbf{0} & \mathbf{0} \\ & \ddots & & \vdots \\ \mathbf{0} & & \mathsf{I}_K & \mathbf{0} \end{bmatrix}$ is a $(Kp \times Kp)$ matrix, $\mathcal{H} = \begin{bmatrix} \mathcal{M} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} = \iota_1 \otimes \mathcal{M}$ is a $(Kp \times [M-1])$ matrix and $J = \begin{bmatrix} \mathsf{I}_K & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} = \iota'_1 \otimes \mathsf{I}_K$ is a $(K \times Kp)$ matrix. Furthermore denote $\bar{\boldsymbol{\mu}} =\mathsf{E}[\mathbf{y_t}]$.

Proposition 3. *Let y_t be an MSI(M)-VAR(p) process as in (5). Then the optimal predictor $\hat{y}_{t+h|t}$ is given by*

$$\hat{y}_{t+h|t} - \mu_y = \underbrace{\left(\sum_{i=1}^{h} J\mathbf{A}^{h-i}\mathcal{H}\mathcal{F}^i\right)}_{K_h}\hat{\zeta}_{t|t} + J\mathbf{A}^h(\mathbf{y}_t - \bar{\boldsymbol{\mu}}). \tag{28}$$

Proof. The optimal predictors results as solution of the following system of linear difference equations:

$$\begin{aligned} \hat{\mathbf{y}}_{t+h|t} - \bar{\boldsymbol{\mu}} &= \mathcal{H}\hat{\zeta}_{t+h|t} + \mathbf{A}(\hat{\mathbf{y}}_{t+h-1|t} - \bar{\boldsymbol{\mu}}) && (29)\\ \hat{\zeta}_{t+h|t} &= \mathcal{F}^h\hat{\zeta}_{t|t}, && (30) \end{aligned}$$

or in a compact notation:

$$\hat{\mathbf{y}}_{t+h|t} = \bar{\boldsymbol{\mu}} + \sum_{i=1}^{h}\mathbf{A}^{h-i}\mathcal{H}\hat{\zeta}_{t+h-i|t} + \mathbf{A}^h(\mathbf{y}_t - \bar{\boldsymbol{\mu}}). \tag{31}$$

Thus, the desired predictor $\hat{y}_{t+h|t} = J\hat{\mathbf{y}}_{t+h|t}$ is given by the solution of the linear difference equation system (29)/(30). ■

In contrast to linear VAR(p) models, the optimal predictor $\hat{y}_{t+h|t}$ depends not only on the last p observations $\mathbf{y}_t$, but is based on the full sample information Y_t through $\hat{\xi}_{t|t}$. This implies a **dynamic** intercept correction $K_h\widehat{\zeta}_{t|t}$ which depends on the persistence of the regimes: $K_h \to \mathbf{0}$. The optimal forecasting rule becomes linear in the limit as the regimes become completely unpredictable:

Corollary 1. *Let y_t be an MSI(M)-VAR(p) process, $p > 0$, as in (6). If the regime s is Granger non-causal or the observed times series vector y (in a* **strict** *sense) iff the regime is unpredictable, $\hat{\xi}_{t+h|t} = \bar{\xi}$. Under this condition, the MSI-VAR model is observationally equivalent to a* **linear** *VAR model with heteroscedastic, non-Gaussian errors (mixture of normals).*

Proof. Unpredictability of the regime-generating process s_t implies that $\mathcal{F}^h\hat{\boldsymbol{\zeta}}_{t|t} = \mathbf{0}_{M-1,1}$. From proposition 3 follows that under this condition the optimal predictor $\hat{y}_{t+h|t}$ is given by

$$\hat{y}_{t+h|t} - \mu_y = J\mathbf{A}^h(\mathbf{y}_t - \bar{\boldsymbol{\mu}})$$

Henceforth, $\hat{\boldsymbol{\zeta}}_{t|t}$ is Granger non-causal for the observed times series vector y_t. ■

Although the optimal predictor is linear in the vector of filtered regime probabilities $\hat{\xi}_{t|t}$ and the last p observations of Y_t, $\hat{y}_{t+h|t}$ is a non-linear function of the observed Y_t as the regime inference $\hat{\xi}_{t|t}$ depends on Y_t in a non-linear fashion. This is illustrated in the following simple example of a univariate two-regime first-order-autoregressive MSI model:

Example 2. *Let y_t denote an MSI(2)-AR(1) model, $y_t = \bar{\nu} + (\nu_1 - \nu_2)\zeta_t + \alpha_1 y_{t-1} + u_t$ with $u_t \sim \mathsf{NID}(0, \sigma_u^2)$. Using the AR(1) representation of the two-state hidden Markov chain, $\zeta_t = \rho\zeta_{t-1} + v_t$ with $\rho = p_{11} + p_{22} - 1$; we get as the optimal h-step predictor:*

$$\hat{y}_{t+h|t} - \bar{\mu} = \alpha^h(y_t - \bar{\mu}) + (\nu_1 - \nu_2)\left(\sum_{i=1}^{h}\alpha^{h-i}\rho^i\right)\widehat{\zeta}_{t|t}. \tag{32}$$

The dynamic intercept correction $\left(\sum_{i=1}^{h}\alpha^{h-i}\rho^i\right)$ in (32) is plotted in the figures 2 and 5 for forecast horizons $h = 1, 2, 4, 8, 12$. For $h = 1$ the corrections depends on the persistence of the regime variable, ρ, but afterwards on the interaction with of dynamics of the regimes with the autoregressive dynamics. This implies for high values of ρ and α, that intercept correction is greater that one, i.e. the impact effect the (reconstructed) regime had on the last observation, $(\nu_1 - \nu_2)\widehat{\zeta}_{t|t}$. Obviously the regime is Granger non-causal if $\rho = 0$.

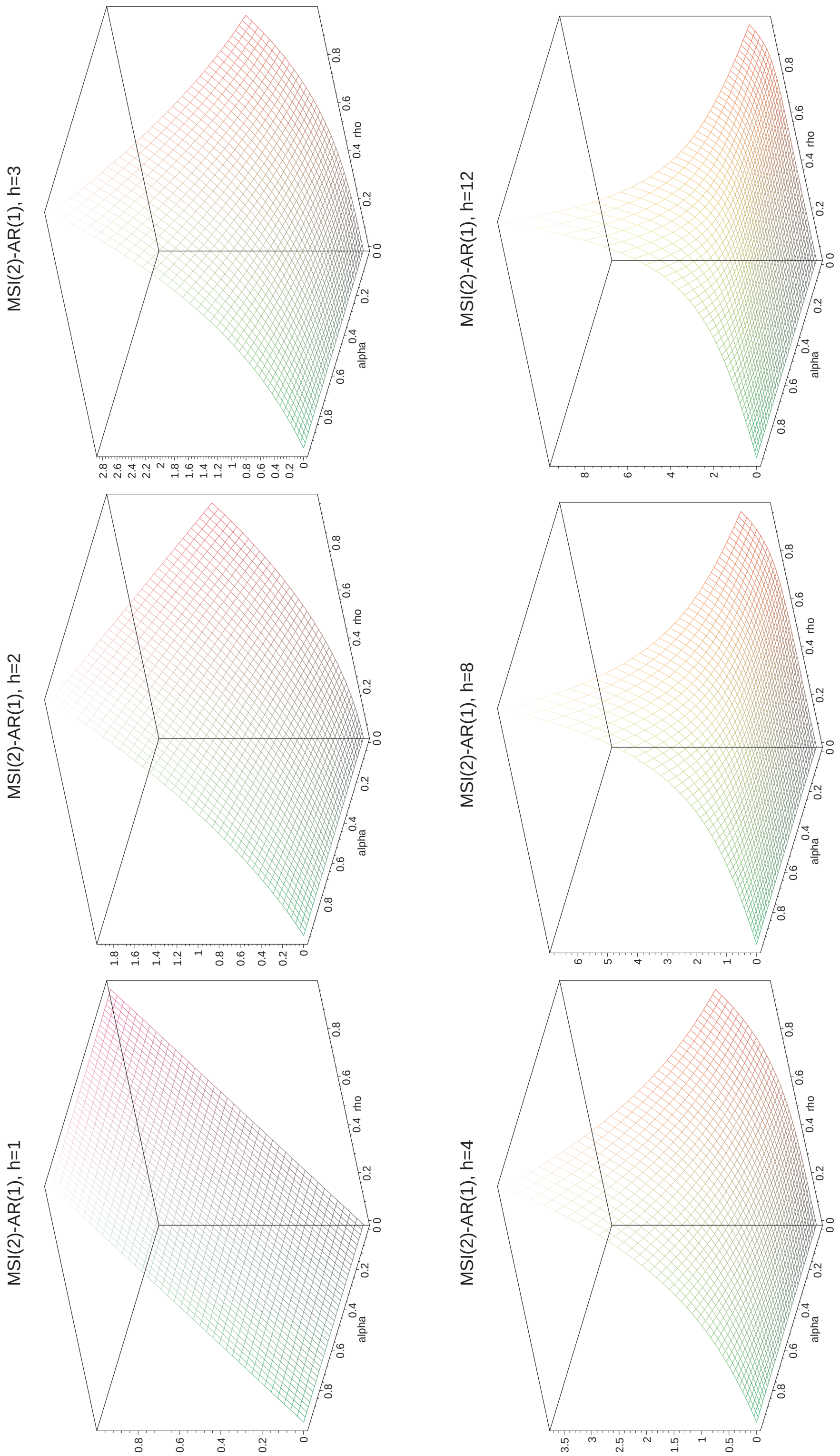

Figure 2   Dynamic Intercept Correction: MSI(2)-AR(1) with $0 < \alpha, \rho < 1$.

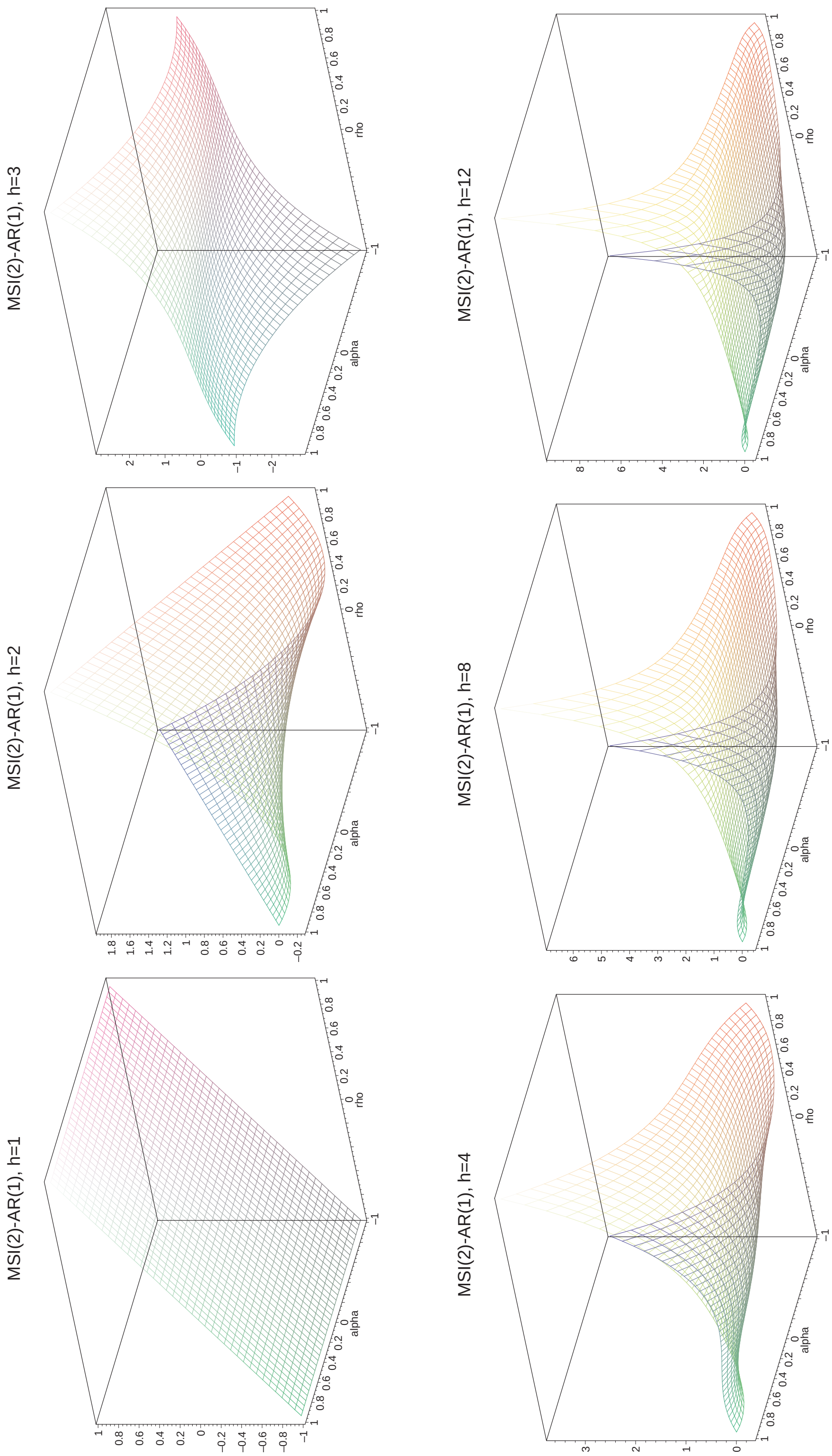

Figure 3 Dynamic Intercept Correction: MSI(2)-AR(1) with $0 < |\alpha|, |\rho| < 1$.

4.2 MSM–VAR Processes

Predictions of MSM(M)-VAR(p) processes can be based on their state-space representation:

$$
\begin{aligned}
y_t - \mu_y &= \mathcal{M}\zeta_t + J_{K,Kp}\mathbf{z}_t, & (33)\\
\mathbf{z}_{t+1} &= \mathbf{A}\mathbf{z}_t + \mathbf{u}_{t+1}, & (34)\\
\zeta_{t+1} &= \mathcal{F}\zeta_t + v_{t+1} & (35)
\end{aligned}
$$

where μ_y is the unconditional mean of y_t, $\mathcal{M} = \begin{bmatrix} \mu_1 - \mu_M & \cdots & \mu_{M-1} - \mu_M \end{bmatrix}$ is $(K \times [M-1])$, $\underline{\mathcal{M}} = \mathbf{I}_P \otimes \mathcal{M}$, $J_{K,Kp} = \iota_1' \otimes \mathbf{I}_K = \begin{bmatrix} \mathbf{I}_K & \mathbf{0}_K & \cdots & \mathbf{0}_K \end{bmatrix}$, ι_1' is the first row of a $(K \times K)$ unit matrix, $\mathbf{z}_t = \mathbf{y}_t - \boldsymbol{\mu}_y - \underline{\mathcal{M}}\zeta_t$
$\begin{bmatrix} y_t - \mu_y - \mathcal{M}\zeta_t \\ y_{t-1} - \mu_y - \mathcal{M}\zeta_{t-1} \\ \vdots \\ y_{t-p+1} - \mu_y - \mathcal{M}\zeta_{t-p+1} \end{bmatrix}$, $\mathbf{u}_t = \begin{bmatrix} u_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$, and $\mathbf{A} = \begin{bmatrix} A_1 & \dots & A_{p-1} & A_p \\ \mathbf{I}_K & & \mathbf{0} & \mathbf{0} \\ & \ddots & & \vdots \\ \mathbf{0} & & \mathbf{I}_K & \mathbf{0} \end{bmatrix}$.

The linear state-space representation (33)–(35) implies that the Gaussian autoregressive process and the regime vector $\zeta_t$enter the system additively. Hence the problem of predicting the observed time series vector y_{t+h} can be reduced to calculating the conditional expectation of the Markovian and the Gaussian component of the state vector $(\mathbf{z}_t', \xi_t{}')'$.

Proposition 4. *Let y_t be an MSM(M)-VAR(p) process as in (5). Then the optimal predictor $\hat{y}_{t+h|t}$ is given by*

$$
\begin{aligned}
\hat{y}_{t+h|t} &= \mu_y + \mathcal{M}\mathcal{F}^h J_{M-1,(M-1)p}\hat{\boldsymbol{\zeta}}_{t|t} + J_{K,Kp}\mathbf{A}^h(\mathbf{y}_t - \boldsymbol{\mu}_y - \underline{\mathcal{M}}\hat{\boldsymbol{\zeta}}_{t|t}) \\
&= \mu_y + J_{K,Kp}\mathbf{A}^h(\mathbf{y}_t - \boldsymbol{\mu}_y) + \left(\mathcal{M}\mathcal{F}^h J_{M-1,(M-1)p} - J_{K,Kp}\mathbf{A}^h\underline{\mathcal{M}}\right)\hat{\boldsymbol{\zeta}}_{t|t.}
\end{aligned}
$$

Proof. By using the law of iterated predictions, we first derive the forecast of ζ_{t+h} conditional on ζ_t and of $\mathbf{z}_{t+h}$ conditional on $\mathbf{z}_t$ respectively:

$$
\begin{aligned}
\mathsf{E}[\mathbf{z}_{t+h}|\mathbf{z}_t] &= \mathbf{A}^j\mathbf{z}_t, \\
\mathsf{E}[\zeta_{t+h}|\zeta_t] &= \mathcal{F}^h\zeta_t.
\end{aligned}
$$

Then, the expectation operator is again applied to the just derived expressions, but now conditional on the sample information Y_t

$$
\begin{aligned}
\hat{\mathbf{z}}_{t|t} &= \mathsf{E}[\mathbf{z}_{t+h}|Y_t] = \mathbf{A}^j\mathsf{E}[\mathbf{z}_{t+h}|Y_t] = \mathbf{A}^j\hat{\mathbf{z}}_{t|t}, \\
\hat{\zeta}_{t+h|t} &= \mathsf{E}[\zeta_{t+h}|Y_t] = \mathcal{F}^h\mathsf{E}[\zeta_{t+h}|Y_t] = \mathcal{F}^h\hat{\zeta}_{t|t},
\end{aligned}
$$

where the filtered Gaussian component $\hat{\mathbf{z}}_{t|t}$ is delivered as a by-product of the filtering procedures for the regimes $\hat{\mathbf{z}}_{t|t} = \mathbf{y}_t - \boldsymbol{\mu}_y - \underline{\mathcal{M}}\hat{\boldsymbol{\zeta}}_{t|t}$. The optimal predictor of the observed time series vector y_{t+h} is the sum of the conditional expectation of the Markovian and the Gaussian states:

$$
\hat{y}_{t+h|t} - \mu_y = \mathcal{M}\mathsf{E}[\zeta_{t+h}|Y_t] + J\mathsf{E}[\mathbf{z}_{t+h}|Y_t] = \mathcal{M}\mathcal{F}^h\hat{\zeta}_{t|t} + J_{K,Kp}\mathbf{A}^j\hat{\mathbf{z}}_{t|t}.
$$

■

It needs no further clarification to verify that for $h \to \infty$ the forecasts $\hat{y}_{t+h|t}$ converge to the unconditional mean of y, if the eigenvalues of $\mathcal{F}$ and $\mathbf{A}$ are inside the unit cycle. In general we find that the following result holds:

Corollary 2. *Let y_t be an MSM(M)-VAR(p) process, $p > 0$, as in (5). Then the unpredictability of the regime s_t does not imply weak Granger non-causality for the observed times series vector y_t.*

Proof. Unpredictability of the regime-generating process s_t implies that $\mathcal{F}^h J_{K,Mp}\, \hat{\boldsymbol{\zeta}}_{t|t} = \mathbf{0}_{M-1,1}$ for $h \geq 1$. From proposition 4 follows that under this condition the optimal predictor $\hat{y}_{t+h|t}$ is given by

$$\hat{y}_{t+h|t} = \mu_y + J_{K,Kp}\mathbf{A}^h(\mathbf{y}_t - \boldsymbol{\mu}_y - \mathcal{M}\hat{\boldsymbol{\zeta}}_{t|t}) \tag{36}$$

Henceforth, $\hat{\boldsymbol{\zeta}}_{t|t}$ is Granger causal for the observed times series vector y_t iff there exists an h with $J_{K,Kp}\mathbf{A}^h\mathcal{M} \neq 0_{K,M-1}$ which follows from the definition of an MSM(M)-VAR(p) process as $\mathbf{A}^h \neq \mathbf{0}_{K,K}$ and $\mathcal{M} \neq \mathbf{0}_{K,M-1}$. ■

We now illustrate proposition 4 with three examples. The first simplifies the previous result by assuming that the lag-order of the VAR is equal to one:

Example 3. *Suppose y_t is an MSM(M)-VAR(1) process. Then the optimal predictor $\hat{y}_{t+h|t}$ is given by*

$$\begin{aligned} \hat{y}_{t+h|t} &= \mu_y + \mathcal{M}\mathcal{F}^h\hat{\zeta}_{t|t} + A^h(y_t - \mu_y - \mathcal{M}\hat{\zeta}_{t|t}) \\ &= \mu_y + A^h(y_t - \mu_y) + \left(\mathcal{M}\mathcal{F}^h - A^h\mathcal{M}\right)\hat{\zeta}_{t|t}. \end{aligned}$$

Further simplifications result if the number of regimes is two which results in the scalar regime generating process discussed before:

Example 4. *Suppose y_t is an MSM(2)-VAR(1) process. Then the optimal predictor $\hat{y}_{t+h|t}$ is given by*

$$\begin{aligned} \begin{bmatrix} \hat{y}_{1,t+h|t}-\mu_{y_1} \\ \hat{y}_{2,t+h|t}-\mu_{y_2} \end{bmatrix} &= \begin{bmatrix} \mu_{11}-\mu_{12} \\ \mu_{21}-\mu_{22} \end{bmatrix}(p_{11}-p_{21})^h\hat{\zeta}_{t|t} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^h \left(\begin{bmatrix} y_{1t}-\mu_{y_1} \\ y_{2t}-\mu_{y_2} \end{bmatrix} - \begin{bmatrix} \mu_{11}-\mu_{12} \\ \mu_{21}-\mu_{22} \end{bmatrix}\hat{\zeta}_{t|t} \right) \\ &= \begin{bmatrix} a_{h,11} & a_{h,12} \\ a_{h,21} & a_{h,22} \end{bmatrix} \begin{bmatrix} y_{1t}-\mu_{y_1} \\ y_{2t}-\mu_{y_2} \end{bmatrix} + \begin{bmatrix} (\mu_{11}-\mu_{12})\left(\rho^h - a_{h,11}\right) + (\mu_{21}-\mu_{22})a_{h,12} \\ (\mu_{21}-\mu_{22})\left(\rho^h - a_{h,22}\right) + (\mu_{11}-\mu_{12})a_{h,21} \end{bmatrix}\hat{\zeta}_{t|t} \end{aligned}$$

where $\rho = p_{11} + p_{22} - 1$ reflects the persistence of the regime-generating process.

Finally, consider the following example of a scalar MSM-AR process:

Example 5. *Consider the MSM(2)-AR(1) process, $y_t - \mu(s_t) = \alpha\,(y_{t-1} - \mu(s_{t-1})) + u_t,\quad u_t \sim \mathsf{NID}(0,\sigma^2)$,which can be rewritten as the sum of two independent processes: $y_t - \mu_y = \mu_t + z_t$,where μ_y is the unconditional mean of y_t, such that $\mathsf{E}[\mu_t] = \mathsf{E}[z_t] = 0$. While the process $z_t = \alpha z_{t-1} + u_t$ is Gaussian, the other component represents the contribution of the Markov chain: $\mu_t = (\mu_1 - \mu_2)\zeta_t$ with $\zeta_t = \xi_{1t} - \bar{\xi}_1$ being $1 - \bar{\xi}_1$ if the regime is 1 and $-\bar{\xi}_1$ otherwise. $\bar{\xi}_1 = p_{21}/(p_{12} + p_{21})$ is the unconditional probability of regime 1 determining the asymmetry of the process; and $\rho = p_{11} + p_{22} - 1$ indicates its persistence. Invoking (12), predictions of the Markov chain are given by $\hat{\zeta}_{t+h|t} = \rho^h\hat{\zeta}_{t|t}$. Thus the optimal predictor $\widehat{y}_{t+h|t} = \mu_y + \hat{\mu}_{t+h|t} + \hat{z}_{t+h|t}$ is given by:*

$$\widehat{y}_{t+h|t} = \mu_y + (\mu_1 - \mu_2)\rho^h\hat{\zeta}_{t|t} + \alpha^h\left[y_t - \mu_y - (\mu_1 - \mu_2)\hat{\zeta}_{t|t}\right]$$

which can be rewritten as

$$\widehat{y}_{t+h|t} = \mu_y + \alpha^h\,(y_t - \mu_y) + (\mu_1 - \mu_2)\left[\rho^h - \alpha^h\right]\hat{\zeta}_{t|t}. \tag{37}$$

The first term in (37) is the optimal prediction rule for a linear model. The contribution of the Markov regime-switching structure is given by the term multiplied by $\hat{\zeta}_{t|t}$, where $\hat{\zeta}_{t|t}$ contains the information about the most recent regime at the time the forecast is made. Thus the contribution of the non-linear part of (37) to the overall forecast depends on both the magnitude of the regime shifts, $|\mu_2 - \mu_1|$ relative to σ_u^2, and on the persistence of regime shifts $\rho = p_{11} + p_{22} - 1$ relative to the persistence of the Gaussian process, given by α. This becomes obvious in the figures 4 and 5 which report the size of the dynamic intercept correction $\rho^h - \alpha^h$ in (37) for forecast horizons $h = 1, 2, 4, 8, 12$. Suppose that $\rho = p_{11} + p_{22} - 1 = \alpha$, $\alpha \neq 0$, then the regime is weakly Granger non-causal for the observed times series vector y despite the predictability of the regime variable.

In the next section we illustrate our analysis by exemplifying the derived forecasting techniques and evaluating the forecast performance of Markov-switching models of the US business cycle. We start with Hamilton (1989), the exact model specification that spearheaded the recent interest in MS-AR models. Two methods of analysis are considered: an empirical forecast accuracy comparison of the MS-AR model with linear and nonlinear alternatives, and a Monte Carlo study. We then show that the forecasting performance of the MS model can be hugely improved by allowing for a third 'high-growth' regime, and simultaneously modelling US output and employment growth in an MS-VAR.

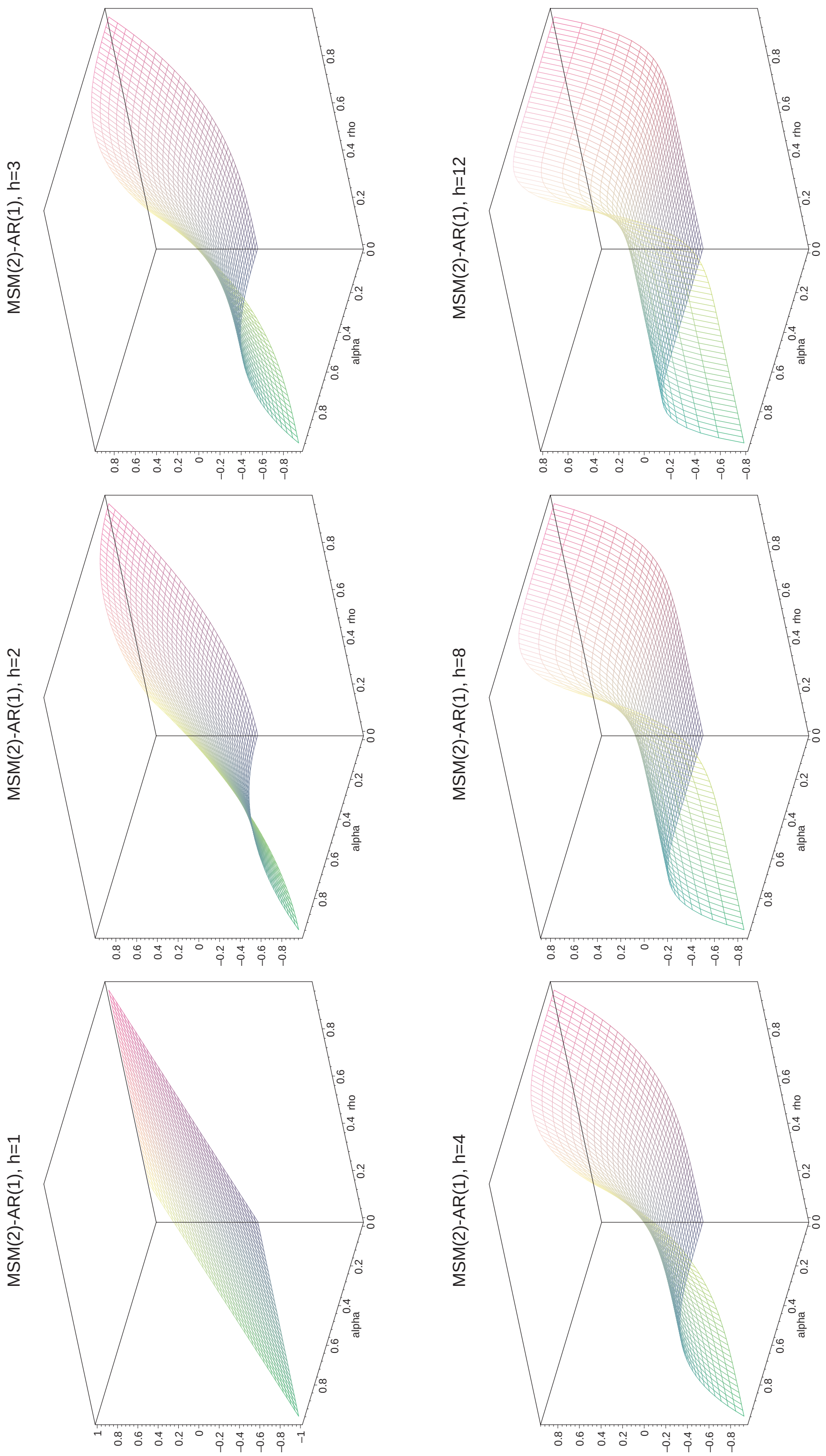

Figure 4 Dynamic Intercept Correction: MSM(2)-AR(1) with $0 < |\alpha|, |\rho| < 1$.

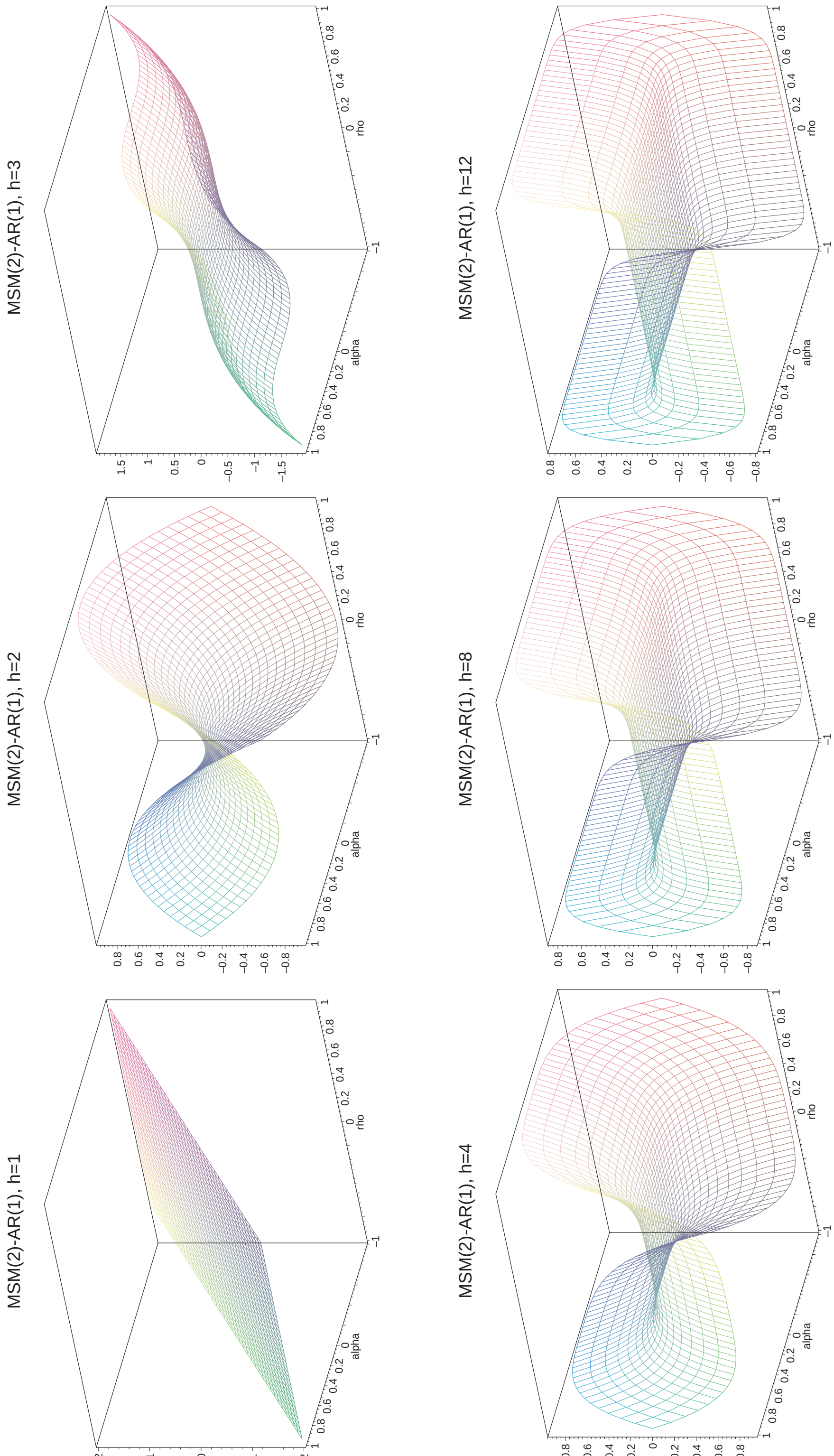

Figure 5 Dynamic Intercept Correction: MSM(2)-AR(1) with $0 < |\alpha|, |\rho| < 1$.

5 Forecasting Economic Time Series with MS-VAR Models – Illustrative Examples

5.1 Hamilton's Model of the US Business Cycle

The Hamilton (1989) model of the US business cycle fostered a great deal of interest in the MS–AR model as an empirical vehicle for characterizing macroeconomic fluctuations, and there have been a number of subsequent extensions and refinements (see the literature discussed in Krolzig, 1997). The Hamilton (1989) model of the US business cycle is an MSM(2)-AR(4) of the quarterly percentage change in US real GNP from 1953 to 1984:

$$\Delta y_t - \mu(s_t) = \alpha_1 \left(\Delta y_{t-1} - \mu(s_{t-1})\right) + \ldots + \alpha_4 \left(\Delta y_{t-p} - \mu(s_{t-4})\right) + u_t, \tag{38}$$

where $u_t \sim \mathsf{NID}(0, \sigma^2)$, and the conditional mean $\mu(s_t)$ switches between two states:

$$\mu(s_t) = \begin{cases} \mu_1 < 0 & \text{if } s_t = 1 \text{ ('contraction' or 'recession')}, \\ \mu_2 > 0 & \text{if } s_t = 2 \text{ ('expansion' or 'boom')}. \end{cases}$$

The variance of the disturbance term, σ^2, is assumed to be the same in both regimes. Thus, contractions and expansions are modeled as switching regimes of the stochastic process generating the growth rate of real GNP. The transition probabilities are constant:

$$\begin{aligned} p_{21} &= \Pr(\text{ contraction in } t \mid \text{expansion in } t-1), \\ p_{12} &= \Pr(\text{ expansion in } t \mid \text{contraction in } t-1). \end{aligned}$$

For a given parametric specification, probabilities are assigned to the unobserved regimes 'expansion' and 'contraction' conditional on the available information set which constitute an optimal inference on the latent state of the economy. Regimes reconstructed in this way are crucial for predicting the probability of future recessions.

5.2 Empirical Forecast Accuracy

The empirical study aims to compare to forecast accuracy of the MS-AR model relatively to linear and non-linear alternatives: the linear autoregressive model and the self-exciting threshold autoregressive (SETAR) model. The SETAR model is another popular non-linear extension of the Box and Jenkins (1970) time-series modeling tradition, applied in the literature to modelling US GNP (see Tiao and Tsay, 1994 and Potter, 1993). The SETAR and MS-AR models differ as to how they model the movement between regimes, and thus the changes in the parameter values of the difference equations that govern the series. While in the MS-AR model the movements between regime are unrelated to the past realizations of the process, and result from the unfolding of an unobserved stochastic process, the SETAR model moves between regimes depending on the past realizations of the process.

For the empirical study, each model is formulated and estimated on a sub-sample of the historical data, and its forecasts of the observations held back at the model specification stage are then evaluated. The forecast accuracy comparison is based on series of 'rolling' forecasts. The Markov-switching, self-exciting threshold and linear autoregressive models of US GNP are estimated once and for all on the sample period considered by Hamilton, 1952:2 - 1984:4 (less observations lost at the beginning of the sample from taking lags). The ML estimation of the MS-AR models used the EM algorithm proposed in Hamilton (1990). Then a sequence of 1 to 8-step ahead forecasts is generated. The forecast origin

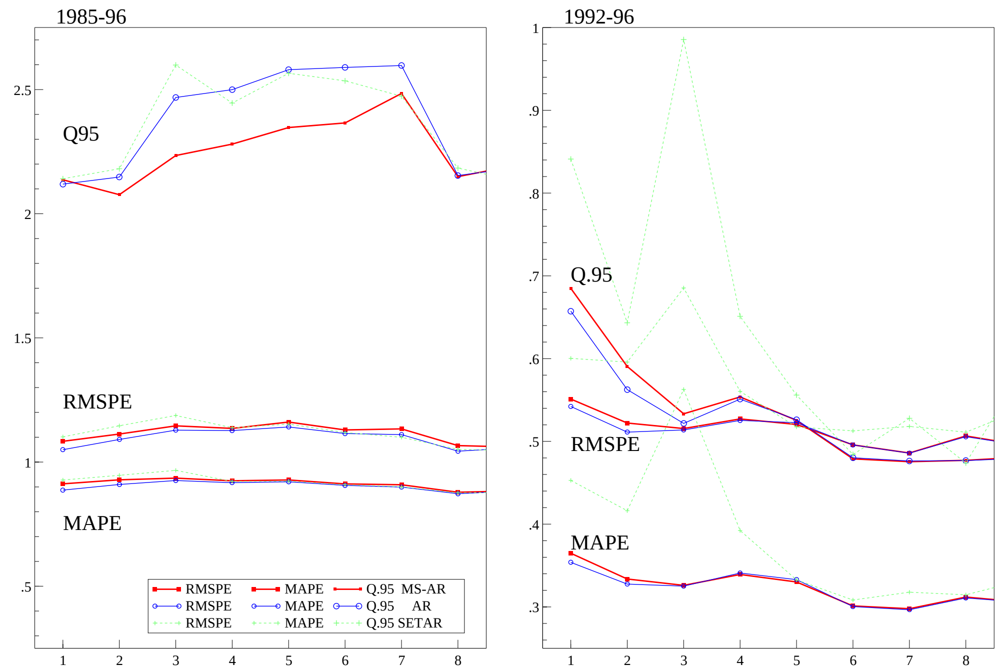

Figure 6 Empirical Forecasting Performance of the Hamilton Model.

is rolled forward one period, and another sequence of 1 to 8-step ahead forecasts is generated. The procedure is then repeated until we have $46\times$ 1-step forecasts, down to $39\times$ 8-step forecasts. This enables root mean squared forecast errors (RMSEs) to be calculated for each forecast horizon. For long horizons the smaller number of forecasts mean that the RMSE calculations are less reliable.

The results of the exercise for the forecast periods 1985:1-1996:2 (succeeding the Hamilton, 1989, sample) and 1992:1-1996:2 are illustrated in figure 6. The lesson to be drawn from these empirical comparisons is that the MS-AR model does not always forecast better. Overall the linear AR model is to be preferred, but records rather small gains. Remarkable is the drop in the forecast uncertainty for the 1992-1996 period.

The failure to improve on the forecast performance of linear models is a finding that warrants further investigation. It may be the case that although non-linearities are a feature of the DGP they are not large enough to yield much of an improvement to forecasting (see Diebold and Nason, 1990). The non-linearities present in the estimation period might not persist in the forecast period due structural breaks occuring after the forecast is made (see Clements and Hendry, 1998). In order to control for factors that might cause the disappointing empirical forecast performance of the MS-AR model, a Monte Carlo (MC) study can be useful.

5.3 Monte Carlo Study

The aim of the Monte Carlo is to learn the systematic gain of the MS model for forecasting purposes. The results of the MC depend on the particular design. In the case of Hamilton (1989), however, the importance of the specificity of the design seems to be a less serious problem as we use an empirical model as the DGP, rather than an artificial DGP whose relevance for actual economics data may be questionable.

We generate data using the Hamilton model as the DGP, then estimate the MS-AR and benchmark models and compare their forecasts. Except for the MSM(2)-AR(4), the linear AR, two-regime MS-AR

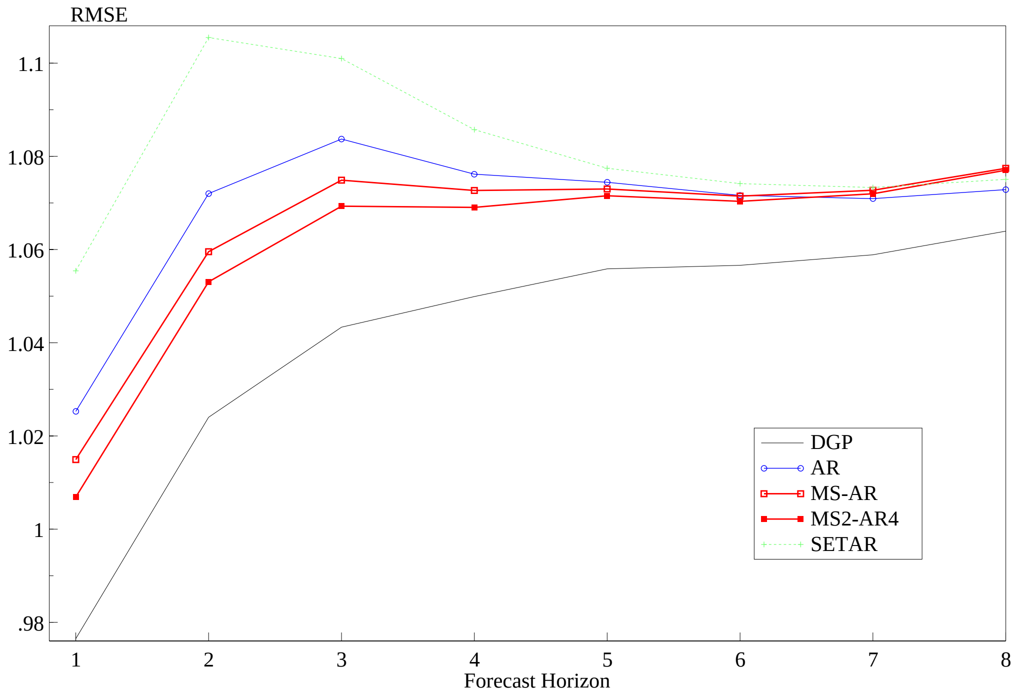

Figure 7 Monte Carlo comparison of the models on RMSE.

and SETAR models are selected data-dependent, so that p (and the delay and threshold for the SETAR) are chosen on each iteration of the Monte Carlo to minimize AIC. The SETAR forecasts are calculated by Monte Carlo using 500 iterations. On a small number of iterations the SETAR model forecasts explode. This only affects longer horizons, and we report RMSEs for the full 1000 replications, but also consider quantiles of the probability distribution of the absolute forecast errors that exclude the errant ones.

The results are summarized in the three figures. Figure 7 shows that the MS-AR is clearly best at short horizons on RMSE when Hamilton's MSM(2)-AR(4) is the DGP. The improvement in the forecast performance is relatively small and occurs only at short horizons. The cost to using the SETAR is greater than those of using the linear AR.

This result is quite independent of the metric used to evaluate the forecast errors, Figure 8 reports RMSEs and mean absolute error (MAE) measures, as well as selected quantiles of the distribution of absolute forecast errors. The legend is as follows: the solid line refers to the DGP, the solid boxes are the MSM(2)-AR(4), the boxes the MSM(2)-AR(p), the circles the AR, and the pluses the SETAR. The Monte Carlo results show that the optimal prediction rule developed in this paper provides a relatively robust forecasting device which can be very close to a linear forecasting rule.

Figure 9 plots the estimated forecast error densities with super-imposed Gaussian densities, for selected horizons. The models have in common that the prediction error densities are clearly non-Gaussian. This implies that a linear Gaussian model is not an preferable forecasting device if forecast intervals are of interest for the researcher despite that the optimal predictors are very close to each other.

Indeed, compared to results of the Monte Carlo, where we controlled for all possible disturbing factors, the empirical forecasting performance of the MS-AR model is disappointing. This suggests that the a structural break occuring after the estimation period might be responsible for this observation. Clements and Krolzig (1998) conclude that the structural instability of two-regime business cycle models is the reason of the little uniformity to the rankings of MS-AR, SETAR and linear models in their empirical forecast accuracy comparison and in their Monte Carlo.

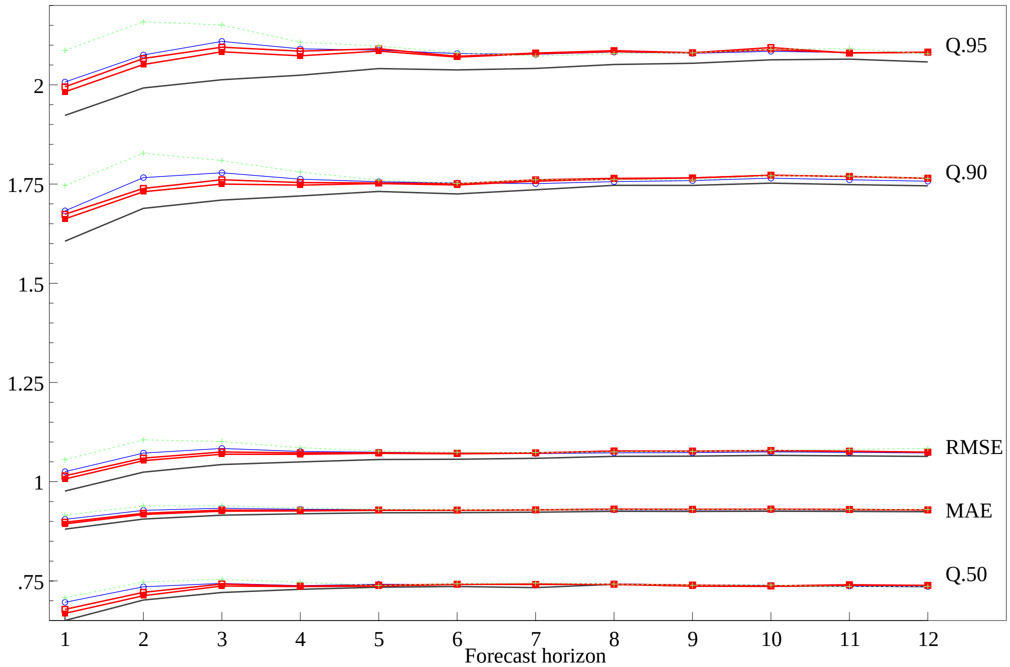

Figure 8 Monte Carlo. Forecast Errors when the DGP is the MSM(2)-AR(4).

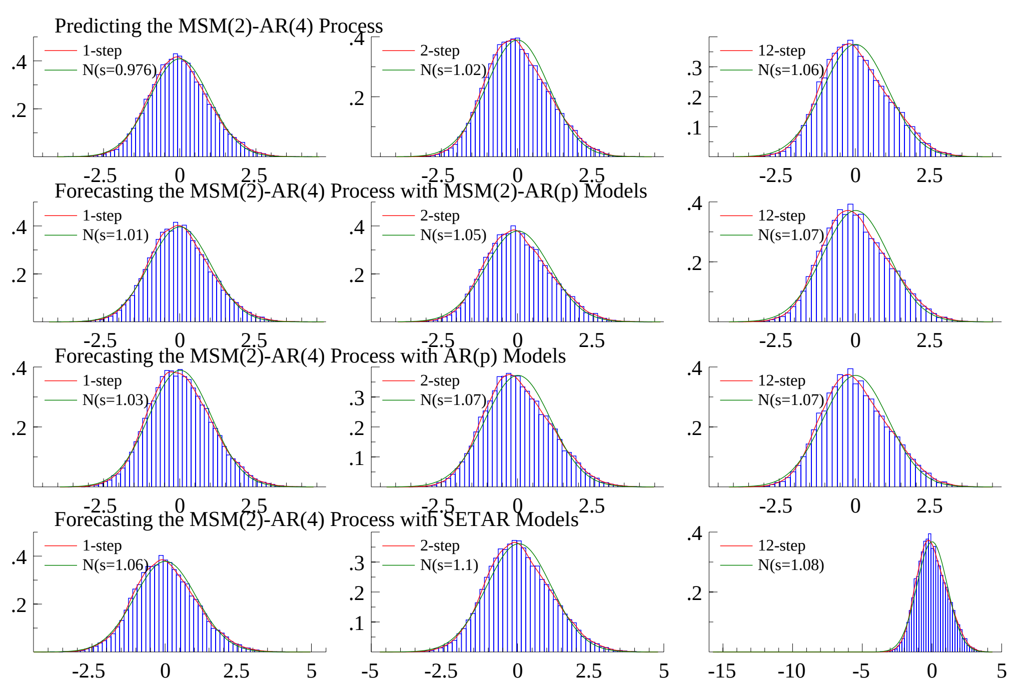

Figure 9 Monte Carlo. Forecast Error Density.

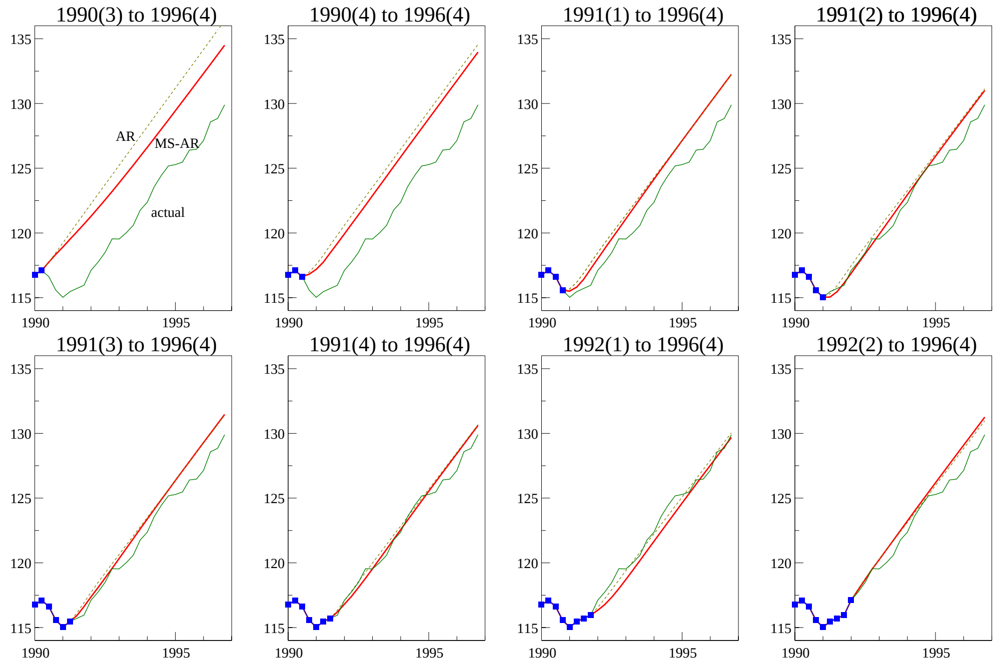

Figure 10 Empirical Forecasting Performance of the Three Regime Model.

We investigate this issue by looking at three-regime models.

5.4 Three-Regime Models of the US Business Cycle

A serious problem associated with the Hamilton model (38) is that for more recent samples (i.e. beyond 1984) Boldin (1996) and Clements and Krolzig (1998) found that an adequate 'business-cycle' model of US GNP (in the sense of generating regime durations consonant with estimates based on the NBER chronology, for example) requires the introduction of a third regime and a regime-dependent error variance. We consider an MSIH(3)-AR(4) model:

$$\Delta y_t = \nu(s_t) + \sum_{j=1}^{4} \alpha_j \Delta y_{t-j} + u_t \tag{39}$$

where $u_t \sim \mathsf{NID}(0, \sigma^2(s_t))$ and α_2 and α_3 being restricted to zero.

Our theory predicts that the differences between the optimal and the linear forecasting rules are strongest if the forecasts are made in the regime which is unlikely to prevail, i.e. during recessions. Therefore we are focusing in our forecasting experiment on the 1990/91 recession. The model are estimated using data until 1990:2 (and further up to 1992:1) and then used for forecasts of Δy_t for the periods 1990:3-1996:4 up to 1992:2-1996:4. Figure 10 gives the cumulated forecasts of the MSIH(3)-AR(4) model (bold line) the nested AR(4) model (dashed line) and compares them with the actual output data. The improvement in the forecasts with the MSIH-AR models over this recessionary period is striking.

It is worth investigating whether the multiple time series approach discussed earlier in this paper is able to improve the forecasts further.

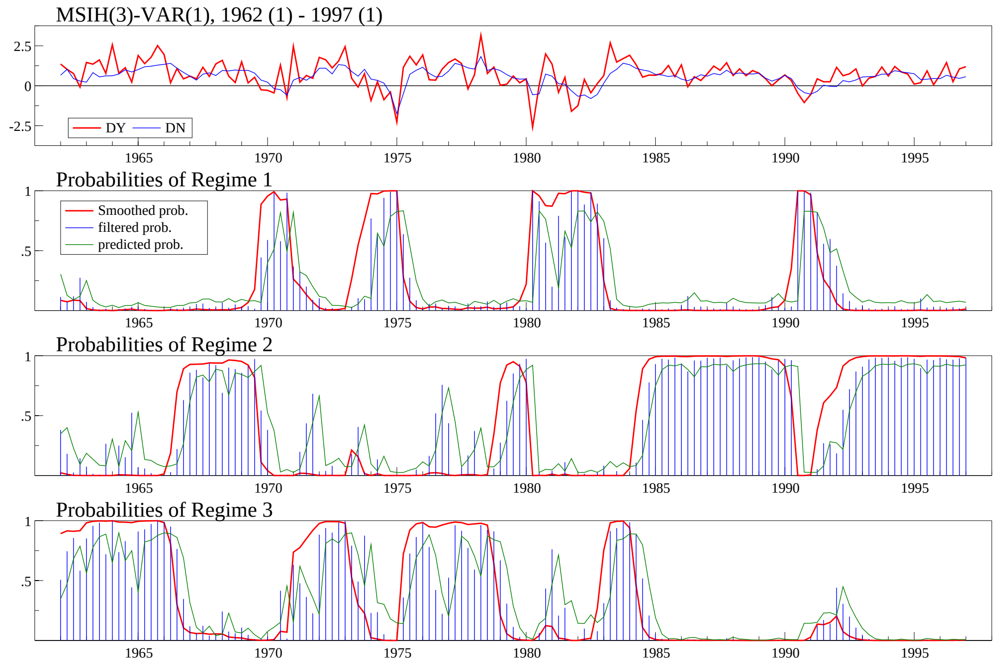

Figure 11 The US Business Cycle.

5.5 Three-Regime Models of US Output and Employment growth Business Cycle

In the following we consider a bivariate, three-regime model of post-war US employment growth, Δn_t, and output growth, Δy_t, in the tradition of Krolzig and Toro (1998). The model is a vector autoregressive Markov-switching process, where some parameters are changing according to the state of common latent regime variable, s_t which represents the phase of the business and employment cycle:

$$\begin{bmatrix} \Delta y_t \\ \Delta n_t \end{bmatrix} = \begin{bmatrix} \nu_1(s_t) \\ \nu_2(s_t) \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta n_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}, \qquad (40)$$

where $u_t \sim \mathsf{NID}(0, \Sigma(s_t))$.

The resulting regime probabilities are plotted in Figure 11. The filtered regime probabilities are shown with a dashed line and the smooth probabilities are shown with a bold line. The filtered probability can be understood as an optimal inference on the state variable (whether the system is in a boom or recession) at time t using only the information up to time t, *i.e.* $\Pr(s_t = m \mid Y_t)$, where m stands for a given regime. The smoothed probability stands for the optimal inference on the regime at time t using the full sample information, $\Pr(s_t = m \mid Y_T)$. It can be seen that regime 1 depicts very precisely the recessions of 1970, 1973/74, 1979/80 and 1990. Regime 2 represents normal growth episodes; while regime 3 characterizes high-growth episodes after recessions. Note that regime 3 is observed until 1985 only, which might indicate a structural change in the phase structure of the business cycle. Expansions after 1985 (regime 2) are characterized by a lower mean growth rate and reduced volatility of macroeconomic fluctuations. This structural break in the volatility of US output growth coincides with the findings of McConnell and Quiros (1998). They found a substantial reduction in the volatility of durable goods production beginning with the first quarter of 1984, which appears to be correlated with a decline in the share of durable goods accounted for by inventories.

Figure 10 reports the results from the replication of the forecasting experiment during the 1990/91 recession. The MSIH(3)-VAR(1) in (40) and the corresponding linear VAR(1) are estimated using

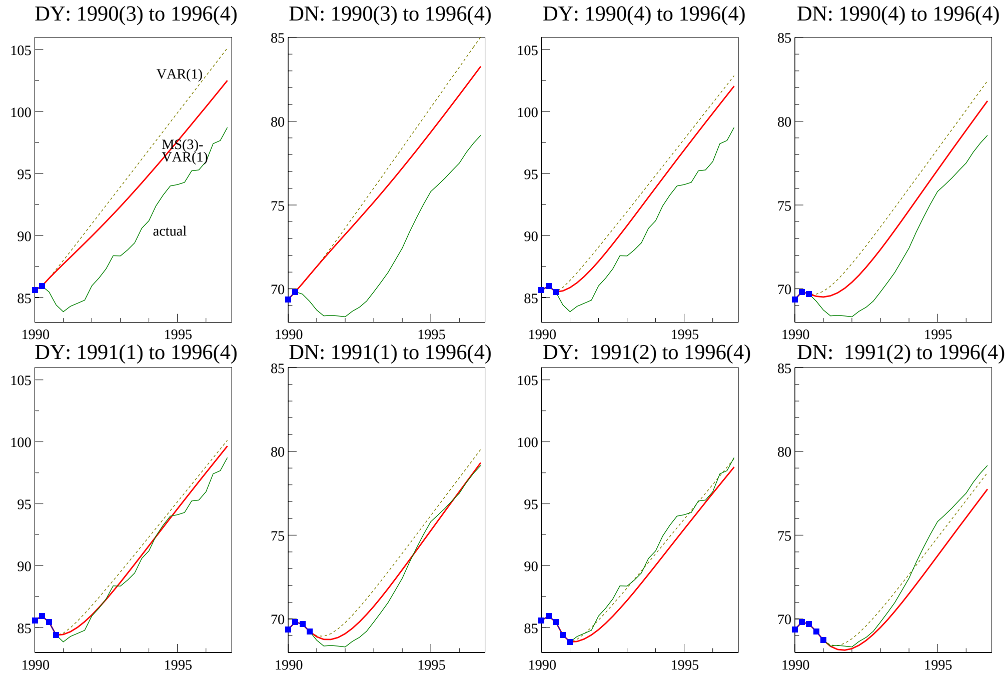

Figure 12 Empirical Forecasting Performance of the US Output and Employment Model.

samples from 1962:1 to 1990:2 (further up to 1991:1). The parametrized models are then used to forecast output and empoyment in the periods 1990:3-1996:4, 1990:4-1996:4 and so on up to 1991:3-1996:4. The MSIH(3)-VAR(1) model (bold line) is again consistently outperforming the linear VAR(1) model (dashed line) provided evidence for an enhanced forecast performance by modelling the regime-switching nature of the DGP with MS-VAR models.

6 Conclusions

One major objective of time series analysis is the creation of suitable models for prediction. In this paper we have developed a general approach to the prediction of multiple time series subject to Markovian shifts in the regime, which is based on the conditional expectation of MS-VAR processes.

In sharp contrast to other non-linear models, an attractive feature of MS-VAR models is the ease with which multi-step forecasts can be obtained when the autoregressive parameters are regime-invariant. This feature allowed us to analyze the implications of the predictability and Granger-causality of regimes on the optimal prediction rule for MS-VAR model. If the autoregressive parameters are regime-invariant, the optimal predictor is linear in the information set and the vector of filtered regime probabilities. However, due to non-linearity of filter, the optimal predictor has generally not the property of being a linear predictor: In general there does not exist a purely linear representation of the optimal predictor in the information set. Conditions have been worked out under which by the conditional expectation coincide with a linear projection. It was shown that this is related to the predictability of the regime-generating Markov chain.

The feasibility of the proposed forecasting technique in empirical research was demonstrated. Employing the Hamilton (1989) model of the US business cycle and its modern contenders, we presented empirical forecast comparisons over three historical periods, and a simulation-based comparison aimed at controlling for certain factors that might influence the outcome of the empirical comparison. While the Monte Carlos provided evidence for an enhanced forecast performance by modelling the regime-switching nature of the DGP, we conjecture that structural breaks in the forecast period made the Hamilton model less powerful in the empirical forecast experiment. It has been shown that models allowing for three regimes deliver superior forecasts during recessions.

In practice some of the assumptions made have to be relaxed. For example, *unknown* parameters might be replaced by unbiased estimators. The present analysis can easily be extended to integrated-cointegrated mechanisms (see Krolzig, 1996). Despite these limitations, the research results provided in this paper let conclude that, for economic time series affected by changes in regime in the stochastic process generating the data, Markov-switching models can yield some improvements compared to the constant-parameter, linear time-series models of the earlier tradition. Our results are also highlighting limitations and possible extensions to forecast Markov-switching vector autoregressive processes which can be summarized to the following points:

(i) Detecting recent regime shifts is essential to predict MS-VAR processes.
(ii) The *predictability of regimes* and their *Granger causality* for the observed time series are critical for the possible contribution of uncovering regime shifts to forecasting.
(iii) MS-VAR can be approximated relatively well with linear models. This is reflected by a special property of MSI(M)-VAR(p) and MSM(M)-VAR(p) processes: their finite-order mixed VARMA(p^*, q^*) representation.
(iv) MS-VAR processes have *short memory* due to the stationary VAR(1) representation of an ergodic Markov chain. The longer the forecast horizon, the better the linear approximation of the optimal predictor and the greater the potential unpredictability of (detrended) economic time series.
(v) *Forecastability* requires the structural stability of MS-VAR models: potential structural breaks in the pattern of regime shifts result in a disappointing forecast performance.

While the exposition was focusing on the predictability of MS-VAR processes as the properties of class of stochastic processes conditional on a given information set, forecasting is a complex process. Forecasting is a process undertaken for a specific purpose, so its evaluation depends on how well it achieves that intent. Consequently, it is extremely difficult to judge the success the forecasting rules proposed in this paper will have in practice, when the data generating mechanism is unknown and possibly subject to structural change unobserved in the estimation period. In this spirit, we have just laid out the basic step in developing a theory of statistical forecasting using formal estimated econometric models in the face of structural change. But we believe that a deep understanding of the statistical properties of MS-VAR processes is fundamental for the use of this class of econometric model in statistical forecasting. For example, the theory developed in this paper enables researchers to predict the potential gain from using the optimal predictor. Parametric conditions have been derived under which simple linear prediction rules are optimal even under presence of regime shifts. The parametric condition for the Granger non-causality of regimes provide a useful, nuisance-free test for the informative value of regime-switching models.

References

Boldin, M. D. (1996). A check on the robustness of Hamilton's Markov switching model approach to the econometric analysis of the business cycle. *Studies in Nonlinear Dynamics and Econometrics*, **1**, 35–46.

Box, G. E. P., and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.

Clements, M. P., and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press. The Marshall Lectures on Economic Forecasting.

Clements, M. P., and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting*. Cambridge, Mass.: MIT Press.

Clements, M. P., and 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., and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*, **13**, 463–475.

Clements, M. P., and Smith, J. (1999). A Monte Carlo study of the forecasting performance of empirical SETAR models. *Journal of Applied Econometrics*, **14**, 124–141.

Dacco, R., and Satchell, S. (1999). Why do regime-switching models forecast so badly?. *Journal of Forecasting*, **18**, 1–16.

Diebold, F. X., and Nason, J. A. (1990). Nonparametric exchange rate prediction. *Journal of International Economics*, **28**, 315–332.

Diebold, F. X., Ohanian, L. E., and Berkowitz, J. (1994). Dynamic equilibrium economies: A framework for comparing models and data. Discussion Paper, University of Pennsylvania.

Diebold, F. X., Rudebusch, G. D., and Sichel, D. E. (1993). Further evidence on business cycle duration dependence. In Stock, J., and Watson, M. (eds.), *Business Cycles, Indicators, and Forecasting*, pp. 255–280. Chicago: University of Chicago Press and NBER.

Doornik, J. A. (1998). *Object-Oriented Matrix Programming using Ox 2.0*. London: Timberlake Consultants Press.

Durland, J. M., and McCurdy, T. H. (1994). Duration dependent transitions in a Markov model of U.S. GNP growth. *Journal of Business and Economic Statistics*, **12**, 279–288.

Filardo, A. J. (1994). Business–cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics*, **12**, 299–308.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica*, **37**, 424–438.

Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime. An investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control*, **12**, 385–423.

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

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

Hendry, D. F. (1999). On detectable and non-detectable structural change. Mimeo, Nuffield College, University of Oxford.

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

Krolzig, H.-M. (1996). Statistical analysis of cointegrated VAR processes with Markovian regime shifts. SFB 373 Discussion Paper 25/1996, Humboldt Universität zu Berlin.

Krolzig, H.-M. (1997). *Markov Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis*. Berlin: Springer.

Krolzig, H.-M. (1998a). Econometric modelling of Markov-switching vector autoregressions using MSVAR for Ox. Discussion Paper, Department of Economics, University of Oxford: http://www.economics.ox.ac.uk/hendry/krolzig.

Krolzig, H.-M. (1998b). Multi-move Gibbs sampling of Markov-switching vector autoregressive processes. Discussion Paper, Department of Economics, University of Oxford.

Krolzig, H.-M., and Toro, J. (1998). A new approach to the analysis of shocks and the cycle in a model of output and employment. Working paper eco 99/30, EUI, Florence.

Lahiri, K., and Wang, J. G. (1994). Predicting cyclical turning points with leading index in a Markov switching model. *Journal of Forecasting*, **13**, 245–263.

McConnell, M. M., and Quiros, G. P. (1998). Output fluctuations in the united states: what has changed since the early 1980s. Staff report no. 41, Federal Reserve Bank of New York.

Pesaran, M. H., and Potter, S. M. (1997). A floor and ceiling model of US Output. *Journal of Economic Dynamics and Control*, **21**, 661–695.

Potter, S. M. (1993). A nonlinear approach to US GNP. *Journal of Applied Econometrics*, **10**, 109–125.

Tiao, G. C., and Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, **13**, 109–131.

Dr Hans-Martin Krolzig is Research Fellow of Nuffield College and Research Officer at the Department of Economics, University of Oxford. His research focuses on econometric methodology, regime-switching models and the analysis of the business cycle.