Forecasts with single-equation Markov-switching model: an application to the gross domestic product of Latvia

Bušs, Ginters
Riga Technical University, Central Statistical Bureau of Latvia

14. February 2010

Forecasts with single-equation Markov-switching model: an application to the gross domestic product of Latvia *

Ginters Buss†
Riga Technical University and
Central Statistical Bureau of Latvia

February 14, 2010

Abstract

The paper compares one-period ahead forecasting performance of linear vector-autoregressive (VAR) models and single-equation Markov-switching (MS) models for two cases: when leading information is available and when it is not. The results show that single-equation MS models tend to perform slightly better than linear VAR models when no leading information is available. However, if reliable leading information is available, single-equation MS models tend to give somewhat less precise forecasts than linear VAR models.

Keywords: Markov-switching, VAR, forecasting, leading information
JEL code: C13, C22, C32, C51, C52, C53

*Acknowledgments: This work has been supported by the European Social Fund within the project "Support for the implementation of doctoral studies at Riga Technical University". The author is greatly thankful to his supervisor Viktors Ajevskis for his guidance and support.

†The views expressed in this paper are those of the author. No responsibility for them should be attributed to the Central Statistical Bureau of Latvia or Riga Technical University. Address for correspondence: Central Statistical Bureau of Latvia, Lacplesa 1, Riga, LV-1301, Latvia; e-mail: `ginters.buss@rtu.lv`

1 Introduction

To the best of our knowledge, there is no publicly available paper that would discuss the forecasting performance of Markov-switching models used to forecast Latvia's gross domestic product (GDP). Thus, this paper is the first attempt to fill the gap in the forecasting literature by comparing one-period ahead forecasting performance of single-equation Markov-switching (MS) and linear vector-autoregressive (VAR) models using Latvia's macroeconomic data.

We use MS models developed in Hamilton (1989) and in a series of papers by Krolzig (see Krolzig, 1998, 2000, 2003, among others), written in a GAUSS code by Bellone (2005), and adapted to Scilab environment by Dubois and Michaux (2010) in econometrics toolbox Grocer.

The paper is organized as follows. Section 2 describes the model and its estimation. Section 3 presents the results for the one-period ahead forecasting performance of single-equation MS and linear VAR models expressed in terms of root mean squared forecast error (RMSFE) for both cases when leading information is available and when it is not available. Following the results in Buss (2009) that an extra regular differencing of the data might improve the forecasting precision during the switch of the business cycle phases, we also show results for the case when the data are subject to two regular differences, instead of one. Finally, Section 4 concludes.

2 Methodology

2.1 The Model

Consider a single-equation Markov-switching model whose parameters are, at least partly, unconditionally time-varying but constant when conditioned on an unobservable discrete regime variable $s_t \in \{1, \ldots, M\}$:

$$\begin{aligned} y_t &= x_t' \beta_{s_t} + z_t' \delta + u_t \\ u_t | s_t &\sim N(0, \sigma^2_{s_t}), \end{aligned} \tag{1}$$

where y_t is a scalar dependent variable at time t, $x_t = (x_{1t}, \ldots, x_{nt})'$ is an $(n \times 1)$ vector of exogenous regressors at time t subject to switching regimes, $z_t = (z_{1t}, \ldots, z_{qt})'$ is a $(q \times 1)$ vector of exogenous regressors at time t that are not subject to switching regimes, δ is a $(q \times 1)$ vector of regression coefficients which are regime independent, and u_t is a Gaussian error term subject to regime changes.

Parameter shift functions β_{s_t} and $\sigma^2_{s_t}$ describe the dependence of the model's parameters β and σ^2 on the regime variable s_t:

$$\beta_{s_t} = \begin{cases} \beta_1 & \text{if } s_t = 1, \\ \vdots & \\ \beta_M & \text{if } s_t = M, \end{cases} \tag{2}$$

and

$$\sigma^2_{s_t} = \begin{cases} \sigma^2_1 & \text{if } s_t = 1, \\ \vdots & \\ \sigma^2_M & \text{if } s_t = M. \end{cases} \tag{3}$$

Let $\mathcal{I}_{t-1} := (y_{t-1}, \ldots, y_1, x'_{t-1}, \ldots, x'_1, z'_{t-1}, \ldots, z'_1)'$ be the information set at $t-1$. The unobservable realizations of the regime $s_t \in \{1, \ldots, M\}$ are generated by an M-state irreducible ergodic Markov stochastic process that is independent of $\mathcal{I}_{t-1}$ or current exogenous variables, x_t and z_t, and is defined by its transition probabilities:

$$\begin{aligned} p_{ij} &= P(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, x_t, z_t, \mathcal{I}_{t-1}) \\ &= P(s_t = j | s_{t-1} = i), \ \sum_{j=1}^{M} p_{ij} = 1 \text{ for all } i, j \in \{1, \ldots, M\}, \end{aligned} \tag{4}$$

that can conveniently be represented with a transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{M1} \\ p_{12} & \cdots & p_{M2} \\ \vdots & \vdots & \vdots \\ p_{1M} & \cdots & p_{MM} \end{bmatrix}. \tag{5}$$

Model (1) can be reduced to such special cases as

- the mean-variance model:

$$y_t = \beta_{s_t} + u_t,$$

- the Markov-switching autoregressive (MS-AR) regime-dependent model:

$$y_t = \nu_{s_t} + y_{t-1}\beta_{1,s_t} + \cdots + y_{t-n}\beta_{n,s_t} + u_t = (1, y_{t-1}, \ldots, y_{t-n})\beta_{s_t} + u_t,$$

- the MS-AR intercept regime-dependent model:

$$y_t = \nu_{s_t} + y_{t-1}\delta_1 + \cdots + y_{t-q}\delta_q + u_t = \beta_{s_t} + (y_{t-1}, \ldots, y_{t-q})\delta + u_t.$$

with $E[u_t^2 | s_t] = \sigma_{s_t}^2$ or $E[u_t^2 | s_t] = \sigma^2$ for any of the above models.

2.2 Estimation

The model is estimated by the maximum likelihood in the following steps:

1. set the initial values of parameters and estimate the model recursively with the Expectation-Maximization (EM) algorithm,
2. compute smoothed probabilities,
3. forecast observed variables.

The rest of the section gives a more detailed look on the procedure, see also Ch. 22 in Hamilton (1994). Denote ξ_t a random $(M \times 1)$ vector whose jth element is equal to unity if $s_t = j$ and zero otherwise:

$$\xi_t = \begin{cases} (1, 0, 0, \ldots, 0)' & \text{if } s_t = 1, \\ \quad \vdots & \\ (0, 0, 0, \ldots, 1) & \text{if } s_t = M. \end{cases} \tag{6}$$

If $s_t = i$, then the jth element of ξ_{t+1} is a random variable that takes on the value unity with probability p_{ij} and zero otherwise. Thus, the conditional expectation of ξ_{t+1} given $s_t = i$ is

$$E(\xi_{t+1}|s_t = i) = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{iM} \end{bmatrix}, \tag{7}$$

which implies

$$E(\xi_{t+1}|\xi_t) = \mathbf{P}\xi_t$$

and from Markov property

$$E(\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots) = \mathbf{P}\xi_t, \tag{8}$$

which implies

$$\xi_{t+1} = \mathbf{P}\xi_t + v_{t+1}, \tag{9}$$

where

$$v_{t+1} := \xi_{t+1} - E(\xi_{t+1}|\xi_t, \xi_{t-1}, \ldots) \tag{10}$$

is a martingale difference sequence.

Let $\mu_{tj} := E[y_t|s_t = j, x_t, z_t, \mathcal{I}_{t-1}] = x_t'\beta_j + z_t'\delta$ be the conditional expectation of y_t, and let $\alpha := (\beta_j', \delta', \sigma_j)'$ be the vector of parameters characterizing the conditional density. Then the conditional probability density function is

$$f(y_t|s_t = j, x_t, z_t, \mathcal{I}_{t-1}, \alpha) = (2\pi)^{-\frac{1}{2}}\sigma_j^{-1} exp\left(-\frac{(y_t - \mu_{tj})^2}{2\sigma_j^2}\right). \tag{11}$$

If there are M different regimes, then there are M different densities represented by (11), that are collected in an $(M \times 1)$ vector

$$\eta_t = \begin{bmatrix} f(y_t|s_t = 1, x_t, z_t, \mathcal{I}_{t-1}, \alpha) \\ \vdots \\ f(y_t|s_t = M, x_t, z_t, \mathcal{I}_{t-1}, \alpha) \end{bmatrix}. \tag{12}$$

Collect α and the transition probabilities p_{ij} governing (11) in a vector of parameters θ. Let $P(s_t = j|\mathcal{I}_t, \theta)$ denote the analyst's inference about the value of s_t based on data obtained through date t and based on knowledge of the population parameters θ. This inference takes the form of a conditional probability that the analyst assigns to the possibility that the tth observation was generated by regime j. Collect these conditional probabilities $P(s_t = j|\mathcal{I}_t, \theta)$ for $j = 1, 2, \ldots, M$ in an $(M \times 1)$ vector $\hat{\xi}_{t|t}$. Collect the forecasts of such probabilities at $t+1$ given observations obtained through date t, $P(s_{t+1} = j|\mathcal{I}_t, \theta)$, in an $(M \times 1)$ vector $\hat{\xi}_{t+1|t}$.

Note that it is assumed that x_t and z_t are exogenous, that is, x_t and z_t contain no information about s_t beyond that contained in $\mathcal{I}_{t-1}$. Hence, the jth element of $\hat{\xi}_{t|t-1}$ could also be described as $P(s_t = j|x_t, z_t, \mathcal{I}_{t-1}, \theta)$. The jth element of the $(M \times 1)$ vector $(\hat{\xi}_{t|t-1} \odot \eta_t)$, where $\odot$ denotes the element-by-element product, can be interpreted as the conditional joint density-distribution of y_t and s_t:

$$\begin{aligned} &P(s_t = j|x_t, z_t, \mathcal{I}_{t-1}, \theta) \times f(y_t|s_t = j, x_t, z_t, \mathcal{I}_{t-1}, \theta) \\ &\quad = p(y_t, s_t = j|x_t, z_t, \mathcal{I}_{t-1}, \theta). \end{aligned} \tag{13}$$

The density of y_t conditioned on past observables is the sum of the M magnitudes in (13) for $j = 1, 2, \ldots, M$. This sum can be written in vector notation as

$$f(y_t|x_t, z_t, \mathcal{I}_{t-1}, \theta) = \mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t), \tag{14}$$

where $\mathbf{1}$ is an $(M \times 1)$ vector of 1s.

If the joint density-distribution in (13) is divided by the density of y_t in (14), the result is the conditional distribution of s_t:

$$\begin{aligned}\frac{p(y_t, s_t = j|x_t, z_t, \mathcal{I}_{t-1}, \theta)}{f(y_t|x_t, z_t, \mathcal{I}_{t-1}, \theta)} &= P(s_t = j|y_t, x_t, z_t, \mathcal{I}_{t-1}, \theta) \\ &= P(s_t = j|\mathcal{I}_t, \theta).\end{aligned}$$

Hence, from (14),

$$P(s_t = j|\mathcal{I}_t, \theta) = \frac{p(y_t, s_t = j|x_t, z_t, \mathcal{I}_{t-1}, \theta)}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)}. \tag{15}$$

But recall from (13) that the numerator in the expression on the right side of (15) is the jth element of the vector $(\hat{\xi}_{t|t-1} \odot \eta_t)$ while the left side of (15) is the jth element of the vector $\hat{\xi}_{t|t}$. Thus, collecting the equations in (15) for $j = 1, 2, \ldots, M$ into an $(M \times 1)$ vector produces

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \odot \eta_t)}{\mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)}. \tag{16}$$

To obtain $\hat{\xi}_{t+1|t}$, take expectations of (9) conditional on $\mathcal{I}_t$:

$$E(\xi_{t+1}|\mathcal{I}_t) = \mathbf{P}E(\xi_t|\mathcal{I}_t) + E(v_{t+1}|\mathcal{I}_t). \tag{17}$$

Since v_{t+1} is a martingale difference sequence with respect to $\mathcal{I}_t$, (17) becomes

$$\hat{\xi}_{t+1|t} = \mathbf{P}\hat{\xi}_{t|t}. \tag{18}$$

The optimal inference and forecast for each date t in the sample are found by iterating on the pair of equations (16) and (18). Given a starting value $\hat{\xi}_{1|0}$ and an assumed value for the population parameter vector θ, one can iterate on (16) and (18) for $t = 1, 2, \ldots, T$ to calculate the values of $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t+1|t}$ for each date t in the sample. The log likelihood function $\mathscr{L}(\theta)$ for the observed data $\mathcal{I}_T$ evaluated at the value of θ that is used to perform the iterations is calculated as a by-product of this algorithm from

$$\mathscr{L}(\theta) = \sum_{t=1}^{T} \log f(y_t|x_t, z_t, \mathcal{I}_{t-1}, \theta), \tag{19}$$

where

$$f(y_t|x_t, z_t, \mathcal{I}_{t-1}, \theta) = \mathbf{1}'(\hat{\xi}_{t|t-1} \odot \eta_t)$$

and maximized numerically with respect to θ.

When the model is estimated, the smoothed probabilities can be calculated. Let $\hat{\xi}_{t|\tau}$ represent the $M \times 1$ vector whose jth element is $P(s_t = j|\mathcal{I}_\tau, \theta)$. For $t < \tau$ it represents the smoothed inference about the regime the process was in at date t based on data obtained on through some later date τ. Smoothed inferences are calculated using an algorithm developed by Kim(1993). In vector form, this algorithm can be written as

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \{\mathbf{P}'[\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]\}, \tag{20}$$

where the sign $\oslash$ denotes element-by-element division. The smoothed probabilities $\hat{\xi}_{t|T}$ are found by iterating on (20) backward for $t = T-1, T-2, \ldots, 1$. This iteration is started with $\hat{\xi}_{T|T}$, which is obtained from (16) for $t = T$. This algorithm is valid only when s_t follows a first-order Markov chain, when the conditional density (11) depends on $s_t, s_{t-1}, \ldots$ only through the current state s_t, and when x_t and z_t , the vectors of explanatory variables other than the lagged values of y_t, are strictly exogenous, meaning that x_t and z_t are independent of s_τ for all t and τ (see Appendix 22.A in Hamilton, 1994).

From the conditional density (11) it is straightforward to forecast y_{t+1} conditional on knowing $\mathcal{I}_t, x_{t+1}, z_{t+1}$ and s_{t+1}:

$$E(y_{t+1}|s_{t+1} = j, x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta) = x'_{t+1}\beta_j + z'_{t+1}\delta. \tag{21}$$

There are M different conditional forecasts associated with the M possible values for s_{t+1}. Note that the unconditional forecast based on actual observable variables is related to these conditional forecasts by

$$\begin{aligned}
&E(y_{t+1}|x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta) \\
&\quad = \int y_{t+1} f(y_{t+1}|x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta)\, dy_{t+1} \\
&\quad = \int y_{t+1} \left\{ \sum_{j=1}^{M} p(y_{t+1}, s_{t+1} = j|x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta) \right\} dy_{t+1} \\
&\quad = \int y_{t+1} \left\{ \sum_{j=1}^{M} [f(y_{t+1}|s_{t+1} = j, x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta) P(s_{t+1} = j|x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta)] \right\} dy_{t+1} \\
&\quad = \sum_{j=1}^{M} P(s_{t+1} = j|x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta) \int y_{t+1} f(y_{t+1}|s_{t+1} = j, x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta)\, dy_{t+1} \\
&\quad = \sum_{j=1}^{M} P(s_{t+1} = j|\mathcal{I}_t, \theta) E(y_{t+1}|s_{t+1} = j, x_{t+1}, z_{t+1}, \mathcal{I}_t, \theta).
\end{aligned}$$

Note that although the Markov chain itself admits the linear representation, $\xi_{t+1} = \mathbf{P}\xi_t + v_{t+1}$, the optimal forecast of y_{t+1} is a nonlinear function of observables, since the inference $\hat{\xi}_{t|t}$ in (16) depends nonlinearly on $\mathcal{I}_t$.

3 Results

The dependent variable of the model (1) is Latvia's quarterly GDP series from 1995Q1 till 2009Q3. The explanatory variables are an aggregate of a few components of the GDP from the production side (cp), imports (imp), net exports (nx), and money supply M1 (m). All series are quarterly, expressed in logs, and once regularly and once seasonally differenced, except m, that is not seasonally differenced. For comparison, we also run models on series for which one regular differencing of the data is replaced with two regular differences due to the result in Buss (2009) that an extra regular differencing might improve the precision of one-period ahead forecasts during a switch of the business cycle phases. Appendix 1 contains a more detailed description of the data. All calculations are performed in Scilab with the aid of its econometrics toolbox Grocer.

The results are summarized in four tables. Tables 1 and 2 show the results about models imposed on once regularly differenced series, while Tables 3 and 4 - on twice regularly differenced series. Tables 1 and 3 give results about models that use leading information, that is, it is assumed in these tables that variables cp, imp, nx and m are known one period ahead of the GDP, so the four explanatory variables are used as leading indicators. Note that, in this case, the exogeneity assumption of the explanatory variables no longer holds for single-equation MS model. Nevertheless, we can still analyze the model's forecasting performance, bearing in mind that not all of the model's assumptions are satisfied. On the contrary, Tables 2 and 4 assume there is no leading information, so the timing of the four explanatory variables there coincide with the timing of the GDP.

Table 1 shows root mean squared forecast errors (RMSFE) for the full sample, the first half of the sample ($\text{RMSFE}_{1.\text{half}}$) and the second half of the sample ($\text{RMSFE}_{2.\text{half}}$) from one-period ahead pseudo real-time forecasts beginning at sample size 19 from single-equation Markov-switching (MS) and linear vector autoregressive (VAR) models with leading information on once regularly differenced series to compare the predictive performance of the two types of models. VAR models are specified by their endogenous variables (first parenthesis) and a lag order (second parenthesis). MS models are specified by the number of regimes (first parenthesis), whether the model is autoregressive (AR) or autoregressive distributed lag (ARDL) model with the explanatory variables other than the endogenous lagged variable specified in parenthesis; next, follows the specification of the lag orders for each explanatory variable along the indication of their dependence (s) or independence (ns) of the regime; finally, it is specified whether the model contains an intercept (c) and whether the intercept is switching (s) or not (ns). The sample is split in halves because the first half of the sample contains a smooth growth whereas the second half of the sample contains a rapid economic downturn, so one can see how the forecasting performance of the models changes along a business cycle. The least RMSFE for each sample space is framed.

The forecasting performance of VAR models is shown in the first four rows in Table 1. Model (1) contains two endogenous variables, GDP and cp; one can see from models (2) to (4) that an addition of endogenous variable nx to the model slightly deteriorates the one-period ahead forecasting performance of the first model, whereas an addition of imp or m slightly improves it for the second half of the sample.

Now, let us discuss the forecasting performance of MS models, starting with MS-AR ones. Model (5) is the mean-variance model since it does not contain any other regressors than a switching constant. Models (6) and (7) are MS-AR models with switching autoregressive coefficient and a non-switching constant. Model (8) is an MS-AR model with non-switching autoregressive coefficients but a switching intercept. It can be seen that these four models, (5)-(8), are doing

Nº	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(2)	0.0186595	0.0142492	0.0222107
2	VAR(GDP,cp,nx)(2)	0.0198866	0.0169021	0.0224782
3	VAR(GDP,cp,imp)(2)	$\boxed{0.0175402}$	0.0154376	$\boxed{0.0194165}$
4	VAR(GDP,cp,m)(2)	0.0187341	0.0149037	0.0219047
5	MS(2)-AR(0)-c(s)	0.0289542	0.0165902	0.0374361
6	MS(2)-AR(3)(s)-c(ns)	0.0288561	0.0156118	0.0377043
7	MS(2)-AR(2)(s)-c(ns)	0.0284772	0.0164813	0.0367461
8	MS(2)-AR(2)(ns)-c(s)	0.0324838	0.0174618	0.0424910
9	MS(2)-ARDL(cp)(1,2)(s)-c(ns)	0.0176693	$\boxed{0.0127261}$	0.0215047
10	MS(2)-ARDL(cp,nx)(1,2,2)(s)-c(ns)	0.0182237	0.0156688	0.0204620
11	MS(2)-ARDL(cp,imp)(1,2,1)(s)-c(ns)	0.0182519	0.0146824	0.0212295
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0207493	0.0174545	0.0235882
13	MS(2)-ARDL(cp,nx)(2(s),{2,2}(ns))	0.0175944	0.0144578	0.0202508
14	MS(2)-ARDL(cp,imp)(2(s),{2,2}(ns))	0.0176390	0.0149478	0.0199708

Table 1: A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models with leading information on once regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample. The least RMSFE in each sample space is framed.

Nº	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(2)	0.0292156	0.0167268	0.0377798
2	VAR(GDP,cp,nx)(2)	0.0370149	0.0247806	0.0461100
3	VAR(GDP,cp,imp)(2)	0.0341238	0.0196662	0.0440694
4	VAR(GDP,cp,m)(2)	0.0299038	0.0171283	0.0386665
5	MS(2)-AR(0)-c(s)	0.0289542	0.0165902	0.0374361
6	MS(2)-AR(3)(s)-c(ns)	0.0328038	0.0186472	0.0424789
7	MS(2)-AR(2)(s)-c(ns)	$\boxed{0.0282835}$	0.0162663	$\boxed{0.0365420}$
8	MS(2)-AR(2)(ns)-c(s)	0.0366473	0.0181612	0.0485410
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0283754	0.0161356	0.0367419
10	MS(2)-ARDL(cp,nx)(2,2,2)(s)-c(ns)	0.0322806	0.0195304	0.0412630
11	MS(2)-ARDL(cp,imp)(2,2,2)(s)-c(ns)	0.0318800	0.0257724	0.0369926
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0287118	0.0168035	0.0369646
13	MS(2)-ARDL(cp,m)(2(s),{2,2}(ns))	0.0309477	0.0169541	0.0403495
14	MS(2)-ARDL(cp,imp)(2(s),{2,1}(ns))	0.0323059	$\boxed{0.0151246}$	0.0431113

Table 2: A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models without leading information on once regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample. The least RMSFE in each sample space is framed.

Nº	Model	RMSFE	$RMSFE_{1.half}$	$RMSFE_{2.half}$
1	VAR(GDP,cp)(3)	0.0208546	0.0189648	0.0226907
2	VAR(GDP,cp,nx)(3)	0.0224267	0.0219535	0.0229187
3	VAR(GDP,cp,imp)(3)	0.0208160	0.0205227	$\boxed{0.0211232}$
4	VAR(GDP,cp,m)(3)	$\boxed{0.0198595}$	$\boxed{0.0177855}$	0.0218484
5	MS(2)-AR(0)-c(s)	0.0331835	0.0264885	0.0390591
6	MS(2)-AR(3)(s)-c(ns)	0.0266697	0.0222837	0.0306497
7	MS(2)-AR(2)(s)-c(ns)	0.0290660	0.0237741	0.0337925
8	MS(2)-AR(2)(ns)-c(s)	0.0323721	0.0273066	0.0370021
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0252970	0.0200264	0.0298957
10	MS(2)-ARDL(cp)(3,3)(s)-c(ns)	0.0303869	0.0316822	0.0289470
11	MS(2)-ARDL(cp,nx)(2,2,1)(s)-c(ns)	0.0256729	0.0218496	0.0291915
12	MS(2)-ARDL(cp,imp)(2,2,1)(s)-c(ns)	0.0277238	0.0260182	0.0294278
13	MS(2)-ARDL(cp,m)(2,2,1)(s)-c(ns)	0.0275117	0.0248066	0.0301209
14	MS(2)-ARDL(cp,nx)(2(s),{2,1}(ns))	0.0286470	0.0233144	0.0333926
15	MS(2)-ARDL(cp,m)(2(s),{2,2}(ns))	0.0245432	0.0191248	0.0292193
16	MS(2)-ARDL(cp,nx)({3,3}(s),3(ns))	0.0296731	0.0277234	0.0316133
17	MS(2)-ARDL(cp,imp)(3,3,3)(s)-c(ns)	0.0317678	0.0339501	0.0292714
18	MS(2)-ARDL(cp,m)(3,3,3)(s)-c(ns)	0.0330116	0.0335336	0.0324477

Table 3: A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models with leading information on twice regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample. The least RMSFE in each sample space is framed.

Nº	Model	RMSFE	RMSFE1.half	RMSFE2.half
1	VAR(GDP,cp)(3)	0.0283351	0.0251617	0.0313568
2	VAR(GDP,cp,nx)(3)	0.0374566	0.0377379	0.0371554
3	VAR(GDP,cp,imp)(3)	0.0351003	0.0318312	0.0382690
4	VAR(GDP,cp,m)(3)	0.0341215	0.0346078	0.0335972
5	MS(2)-AR(0)-c(s)	0.0331835	0.0264885	0.0390591
6	MS(2)-AR(3)(s)-c(ns)	0.0364793	0.0290636	0.0429787
7	MS(2)-AR(2)(s)-c(ns)	0.0289120	0.0233798	0.0338124
8	MS(2)-AR(2)(ns)-c(s)	0.0320115	0.0272778	0.0363722
9	MS(2)-ARDL(cp)(2,2)(s)-c(ns)	0.0291961	$\boxed{0.0228417}$	0.0346950
10	MS(2)-ARDL(cp,nx)(2,2,2)(s)-c(ns)	0.0339771	0.0292548	0.0383628
11	MS(2)-ARDL(cp,imp)(2,2,2)(s)-c(ns)	0.0317295	0.0266464	0.0363598
12	MS(2)-ARDL(cp,m)(2,2,2)(s)-c(ns)	0.0308232	0.0234447	0.0370879
13	MS(2)-ARDL(cp,m)(2(s),{2,2}(ns))	0.0297117	0.0261638	0.0330669
14	MS(2)-ARDL(cp,imp)(2(s),{2,1}(ns))	0.0313216	0.0245121	0.0372157
15	MS(2)-ARDL(cp)(3,3)(s)-c(ns)	$\boxed{0.0264927}$	0.0230135	$\boxed{0.0297469}$
16	MS(2)-ARDL(cp,nx)(3,3,2)(s)-c(ns)	0.0341934	0.0350868	0.0332178
17	MS(2)-ARDL(cp,imp)(3,3,3)(s)-c(ns)	0.0358542	0.0351741	0.0365629
18	MS(2)-ARDL(cp,m)(3,3,2)(s)-c(ns)	0.0281086	0.0238231	0.0320399

Table 4: A comparison of one-period ahead pseudo real-time forecasting performance from single-equation MS and VAR models without leading information on twice regularly differenced series in terms of RMSFE for the full sample, first half of the sample and second half of the sample. The least RMSFE in each sample space is framed.

poorly in the second half of the sample, although are still competitive in the first half of the sample. Next, consider four MS-ARDL models, (9)-(12), with switching slope coefficients and a non-switching intercept. It can be seen that model (9), which is a close counterpart to model (1), performs slightly better than model (1) in terms of RMSFE in both halves of the sample. Similarly, model (10), which is a close counterpart to model (2), performs slightly better than the latter in all sample spaces. However, models (11) and (12) seem to perform slightly worse than their VAR counterparts, (3) and (4), respectively. Finally, models (13) and (14) exclude a constant and allow for switching coefficients for lagged GDP and *cp*, but fix other coefficients. One can see that these two models perform well in both halves of the sample. To summarize information in Table 1, the forecasting performance of the two model types is similar, with the least RMSFE for the first half of the sample obtained by an MS model, and the least RMSFE for the second half of the sample and the full sample - by linear VAR.

Next, consider Table 2 that summarizes the forecasting performance of linear VAR and single-equation MS models when no leading information is available. As in the previous table, the first four models are VAR ones with different sets of endogenous variables. Model (5) is the mean-variance model with a switching intercept being the only regressor. Regardless of the simplicity of model (5), it shows a slightly better forecasting performance than any of the VAR models considered in any sample space. Models (6) to (8) introduce non-zero number of lags. It can be seen that introducing two lags of the dependent variable and allowing slope coefficients to be regime-dependent, further improves the forecasting performance of the MS-AR model for all sample spaces. Models (9)-(12) are single-equation MS counterparts to the linear VAR models (1)-(4), respectively. One can see that all four MS models perform better than the respective VAR models for all sample spaces, except model (11) for the first half of the sample. The results for models (13)-(14) show that allowing only the coefficients for lagged dependent variable to switch does not improve the forecasting performance. To summarize Table 2, single-equation MS models tend to perform slightly better than linear VAR models in terms of one-period ahead forecasting performance on once regularly differenced series when no leading information is available.

Following the result of Buss (2009) that, during a switch of the business cycle phases, the short-term forecasting performance might improve if two, instead of one, regular differencing is implemented, Tables 3 and 4 show results for twice regularly differenced data. Table 3 summarizes a comparison of one-period ahead pseudo real-time forecasting performance of single-equation MS and linear VAR models with leading information on twice regularly differenced series. One can see that the performance of MS models lags behind that of VAR counterparts. Table 4 shows the results of single-equation MS and linear VAR models without leading information on twice regularly differenced series. Table 4 shows that, comparing the models with the same variables and lag order, single-equation MS models (15)-(18) tend to give smaller RMSFE than the corresponding linear VAR models for all sample spaces, except for model (17) for the first half of the sample.

The results show that, if leading information is available, a second regular differencing of the data does not seem to improve the forecasting precision during a switch of the business cycle phases. However, if no leading information is used, the second regular differencing appears to improve forecasting precision, which is in line with the results in Buss (2009), where no leading information was used.

4 Conclusions

To the best of our knowledge, this is the first publicly available paper that attempts to evaluate short-term forecasting performance of MS models for Latvia's economy. This paper compares one-period ahead pseudo real-time forecasting performance of single-equation MS models compared to linear VAR models with and without leading information. The results show that when leading information is available, the forecasting performance of single-equation MS models is slightly worse than linear VAR models. On the contrary, if there is no leading information at hand, MS models tend to perform somewhat better in terms of one-period ahead forecasts than their linear VAR counterparts.

The results also show that if leading information is available, a second regular differencing of the data does not appear to improve the forecasting precision during a switch of the business cycle phases. However, if no leading information is used, the second regular differencing appears to improve forecasting precision, which is in line with the results in Buss (2009) where no leading information was used.

References

[1] Bellone, B. (2005),"Classical Estimation of Multivariate Markov-Switching Models using MSVARlib", Econometrics 0508017, EconWPA

[2] Buss, G. (2009), "Comparing Forecasts of Latvia's GDP Using Simple Seasonal ARIMA Models and Direct Versus Indirect Approach", MPRA Paper 16832, University Library of Munich, Germany

[3] Dubois, E. and Michaux, E. (2010), "Grocer 1.41: an econometric toolbox for Scilab", available at http://dubois.ensae.net/grocer.html

[4] Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, Vol. 57, No. 2, 357-384

[5] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Ch. 22, 677-703

[6] Krolzig, H. M. (1998), "Econometric Modelling of Markov-Switching Vector Autoregressions using MSVAR for Ox", Working paper

[7] Krolzig, H. M. (2000), "Predicting Markov-Switching Vector Autoregressive Processes", Working paper

[8] Krolzig, H. M. (2003), "Constructing Turning Point Chronologies with Markov-Switching Vector Autoregressive Models: the Euro-Zone Business Cycle", Working paper

Appendix 1

The list of data used in the paper. All national accounts series are chain-priced as of 2000.

Series	Definition	Source
GDP	Gross domestic product	Central Statistical Bureau of Latvia
C	Output in mining and quarrying industry	Central Statistical Bureau of Latvia
D	Output in manufacturing industry	Central Statistical Bureau of Latvia
E	Output in electricity, gas and water supply industry	Central Statistical Bureau of Latvia
F	Output in construction industry	Central Statistical Bureau of Latvia
H	Output in hotels and restaurants industry	Central Statistical Bureau of Latvia
L	Output in public administration and defense, and compulsory social security industries	Central Statistical Bureau of Latvia
D21	Taxes	Central Statistical Bureau of Latvia
cp	Sum of C,D,E,F,H,L, and D21	Derived by the author
exp	Exports	Central Statistical Bureau of Latvia
imp	Imports	Central Statistical Bureau of Latvia
nx	Net exports, exp-imp	Derived by the author
m	Monetary aggregate M1, quarterly average	Bank of Latvia