Statistics and Probability Letters 90 (2014) 108–113

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Combining a regression model with a multivariate Markov chain in a forecasting problem

Bruno Damásio, João Nicolau *

Universidade de Lisboa, ISEG, CEMAPRE, Portugal

ARTICLE INFO

Article history:
Received 21 November 2013
Received in revised form 17 February 2014
Accepted 27 March 2014
Available online 1 April 2014

Keywords:
Multivariate Markov chain
Higher-order Markov chain
Forecasting

ABSTRACT

This paper proposes a new concept: the usage of Multivariate Markov Chains (MMC) as covariates. Our approach is based on the observation that we can treat possible categorical (or discrete) regressors, whose values are unknown in the forecast period, as an MMC in order to improve the forecast error of a certain dependent variable. Hence, we take advantage of the information about the past state interactions between the MMC categories to forecast the categorical (or discrete) regressors and improve the forecast of the actual dependent variable.

1. Introduction

Consider a simple regime-switching model

$$y_t = \beta x_t + \delta z_t + u_t$$

where z_t is a latent dummy variable that evolves over time according to a homogeneous Markov chain (i.e. $P\left(z_t = i_0 | z_{t-1} = i_1\right), i_0, i_1 = 0, 1$). This model and further refinements have been extensively studied in the literature (see Hamilton (1989)). In some circumstances the z_t variable may be observable, and in this case standard methods of estimation of β and δ apply. However, forecasting y_t may raise some difficulties because z_t (which is assumed to be a random variable) is not observable in the forecasting period (to simplify one assumes that x_t is a dynamic term, e.g. AR(1), or a simple trend). In this case a probabilistic structure is needed for z_t, for example a Markov chain, as in regime-switching models. In this paper we analyze the forecasting problem when the y_t variable depends on $s > 1$ discrete or categorical variables (observable during the estimation period), whose dependencies are governed by a multivariate Markov chain. This approach is new in the literature and the closest model to ours is perhaps the regime-switching one cited above. However, in contrast to regime-switching models which only deal with univariate Markov chains, usually with few states (in most cases with two or three states), given the complexity of the estimation procedures, our model is able to involve many "z_t" variables, with multiple states, thanks to the MTD-probit specification as we explain later on.

To be more precise, this paper considers the forecasting of a time series (y_t) that depends on quantitative variable(s) (x_t) and on s discrete or categorical variables, $(S_{1t}, \ldots, S_{st})$ where S_{jt} $(j = 1, \ldots, s)$ can take on values in the finite set $\{1, 2, \ldots, m\}$. We assume that S_{jt} depends on the previous values of $S_{1t-1}, \ldots, S_{jt-1}, \ldots, S_{st-1}$, and this dependence is well modeled by a first-order MMC. However, S_{jt} can also depend on some explanatory variables lagged over more than one period — our approach may in fact be viewed as a higher-order MMC (e.g. we may take S_{jt-1} as S_{t-j}, and in this case we would have an

* Correspondence to: School of Economics and Management (ISEG), Rua do Quelhas 6, 1200-781 Lisboa, Portugal. Tel.: +351 21 3925876; fax: +351 21 3922782.

E-mail address: nicolau@iseg.utl.pt (J. Nicolau).

http://dx.doi.org/10.1016/j.spl.2014.03.026
0167-7152/

s-order Markov chain). We propose using MMC as covariates in a regression model in order to improve the forecast error of a certain dependent variable, provided it is caused, in the Granger sense, by the MMC. Traditionally, and so far, the published literature only addresses the MMC as an end in itself. Here we take advantage of the information about the past state interactions between the MMC categories to forecast the dependent variable more accurately. As far as we know this forecasting problem has not yet been analyzed in the literature.

To form a regression model relating y_t to the categorical variables, we convert the S_{jt} categories into a set of dummy variables as follows:

$$z_{jkt} = \mathcal{I}_{\{S_{jt}=k\}} \tag{1.1}$$

where $\mathcal{I}_{\{.\}}$ is the indicator function, $\mathcal{I}_{\{S_{jt}=k\}} = 1$ if $S_{jt} = k$ and 0 otherwise. The proposed methodology also supports the event where S_{jt} is a discrete variable with state space $\{1, 2, \ldots, m\}$ (say), in which case no dummy variables are needed.

Let us now assume, without any loss of generality, a linear specification like:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \boldsymbol{z}_t'\boldsymbol{\delta} + u_t \tag{1.2}$$

where:

- $\boldsymbol{x}_t'$ may be a vector of both deterministic and stochastic components, like AR(1) or other $\mathcal{F}_{t-1}$ or $\mathcal{F}_t$ measurable predetermined terms. Here $\mathcal{F}_t$ represents the information available at time t, i.e. the σ-algebra generated by all events up to time t.
- $\boldsymbol{z}_t'$ is a vector of dummy variables z_{kjt}, concerning the MMC, defined in (1.1).
- $\{u_t\}$ is a white noise process mean independent of $\boldsymbol{x}_t'$ and $\boldsymbol{z}_t'$. We do not assume any distribution for u_t.

To forecast y_{t+h} we use the best predictor according to the expected squared forecast error:

$$E\left(y_{t+h}\middle|\, \mathcal{F}_t\right) = E\left(\boldsymbol{x}_{t+h}'\middle|\, \mathcal{F}_t\right)\boldsymbol{\beta} + E\left(\boldsymbol{z}_{t+h}'\middle|\, \mathcal{F}_t\right)\boldsymbol{\delta} \tag{1.3}$$

given the exogeneity of the disturbance term, i.e. $E\left(u_t\middle|\, \mathcal{F}_{t-1}\right) = 0\ \forall t$.

To illustrate, suppose that we have two categorical variables ($s = 2$) and each categorical datum takes on values in the set $\{1, 2, 3\}$, i.e. $m = 3$. Unwinding the vector $\boldsymbol{z}_t'$ and the vector $\boldsymbol{\delta}$ it follows that

$$y_{t+h} = \boldsymbol{x}_{t+h}'\boldsymbol{\beta} + \delta_{11}\mathcal{I}_{\{S_{1t}=1\}} + \delta_{12}\mathcal{I}_{\{S_{1t}=2\}} + \delta_{21}\mathcal{I}_{\{S_{2t}=1\}} + \delta_{22}\mathcal{I}_{\{S_{2t}=2\}} + u_t \tag{1.4}$$

where S_{jt} represents the j-th categorical series of the MMC (notice that the dummy variable trap is avoided with this specification). Since the values of S_{jt+h} are unknown in the forecasting periods, i.e. for $h \geqslant 1$, we explore possible dependencies between S_{jt+h} and past values of S_{1t+h} and S_{2t+h} using an MMC approach, to predict S_{jt+h}, and consequently, y_{t+h}. If both S_{1t} and S_{2t} are discrete variables, the regression equation is simpler:

$$y_{t+h} = \boldsymbol{x}_{t+h}'\boldsymbol{\beta} + \delta_1 S_{1t+h} + \delta_2 S_{2t+h} + u_t. \tag{1.5}$$

From Eqs. (1.4) or (1.5), it is clear that to forecast y_{t+h} one needs to evaluate $P\left(S_{jt+h} = k\middle|\, \mathcal{F}_t\right)$, for $k = 1, 2, \ldots, s$. To keep these expressions simple, we make the following assumptions:

Assumption 1.1. First order MMC.

$$P\left(S_{jt} = k\middle|\, \mathcal{F}_{t-1}\right) = P\left(S_{jt} = k\middle|\, S_{1t-1} = i_1, \ldots, S_{st-1} = i_s\right). \tag{1.6}$$

That is, S_{jt} given $\{S_{1t-1}, \ldots, S_{st-1}\}$ is independent of any other variables in $\mathcal{F}_{t-1}$.

Assumption 1.2. Homogeneous MMC.

We have a homogeneous MMC in the sense that

$$P\left(S_{jt} = k\middle|\, S_{1t-1}, \ldots, S_{st-1}\right) = P\left(S_{jt+h} = k\middle|\, S_{1t+h-1}, \ldots, S_{st+h-1}\right). \tag{1.7}$$

Assumption 1.3. Contemporaneous needless terms.

S_{jt} is independent of $\left\{S_{1t}, \ldots, S_{j-1t}, S_{j+1t}, \ldots, S_{st}\right\}$ given $\{S_{1t-1}, \ldots, S_{st-1}\}$, i.e.

$$\begin{aligned} &P\left(S_{jt} = k\middle|\, S_{1t} = i_1, \ldots, S_{j-1t} = i_{j-1}, S_{j+1t} = i_{j+1}, \ldots, S_{st} = i_s, S_{1t-1}, \ldots, S_{st-1}\right) \\ &\quad = P\left(S_{jt} = k\middle|\, S_{1t-1}, \ldots, S_{st-1}\right). \end{aligned} \tag{1.8}$$

To obtain the forecast of y_{t+h} we need to calculate $E\left(\boldsymbol{x}_{t+h}'\middle|\, \mathcal{F}_t\right)$ and $E\left(\boldsymbol{z}_{t+h}'\middle|\, \mathcal{F}_t\right)$. It is assumed the former expression is known, hence we focus on the latter expression. A generic element of $E\left(\boldsymbol{z}_{t+h}'\middle|\, \mathcal{F}_t\right)$ is $E\left(\boldsymbol{z}_{kj,t+h}\middle|\, \mathcal{F}_t\right)$ which, by Assumption 1.1, can be written as

$$\begin{aligned} E\left(\boldsymbol{z}_{kj,t+h}\middle|\, \mathcal{F}_t\right) &= P\left(\boldsymbol{z}_{kj,t+h} = 1\middle|\, \mathcal{F}_t\right) = P\left(S_{j,t+h} = k\middle|\, \mathcal{F}_t\right) \\ &= P\left(S_{j,t+h} = k\middle|\, S_{1t} = i_1, \ldots, S_{st} = i_s\right). \end{aligned} \tag{1.9}$$

We use the MMC theory to estimate the expression (1.9), which ultimately leads to the expressions $E\left(\boldsymbol{z}_{t+h}'\middle|\, \mathcal{F}_t\right)$ and $E\left(y_{t+h}\middle|\, \mathcal{F}_t\right)$. We briefly cover the main aspects of MMC estimation theory in the next section.

2. Multivariate Markov Chains as regressors: model estimation

In this section we explain our strategy to estimate the parameters defined in Eq. (1.2), $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\delta})$ and the parameters associated with the multivariate Markov chain, which we denote by $\boldsymbol{\eta}$. Let $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\eta})$ be the complete vector of parameters, and B and D the parameter space of $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\delta})$ and $\boldsymbol{\eta}$, respectively. Given the structure of our model and by construction, $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ are variation free (see Engle et al. (1983)), since $(\boldsymbol{\psi}, \boldsymbol{\eta}) \in B \times D$, i.e. $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ are not subject to cross restrictions so that for any specific admissible value in B for $\boldsymbol{\psi}$, $\boldsymbol{\eta}$ can take any value in D. In these circumstances, the conditional distribution of $y_t | \mathbf{S}_t, \mathcal{F}_{t-1}$ depends on $\boldsymbol{\psi}$ only, and the conditional distribution of $\mathbf{S}_t | \mathcal{F}_{t-1}$ depends on $\boldsymbol{\eta}$ only. In this way the joint density of the complete sample can be sequentially factorized as follows:

$$
\begin{aligned}
f\left(y_0, y_1, \ldots, y_n; S_{j0}, S_{j1}, \ldots, S_{jn}; \boldsymbol{\theta}\right) &= \prod_{t=1}^{n} f\left(y_t, \mathbf{S}_t | \mathcal{F}_{t-1}; \boldsymbol{\theta}\right) \\
&= \prod_{t=1}^{n} f\left(y_t | \mathbf{S}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}\right) \prod_{t=1}^{n} P\left(\mathbf{S}_t | \mathcal{F}_{t-1}; \boldsymbol{\eta}\right). \qquad (2.1)
\end{aligned}
$$

Let us focus on $P\left(\mathbf{S}_t | \mathcal{F}_{t-1}; \boldsymbol{\eta}\right) = P\left(S_{1t}, \ldots, S_{st} | \mathcal{F}_{t-1}; \boldsymbol{\eta}\right)$. This expression may be written as:

$$
\begin{aligned}
P\left(S_{1t}, \ldots, S_{st} | \mathcal{F}_{t-1}; \boldsymbol{\eta}\right) &= P\left(S_{1t}, \ldots, S_{st} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}\right) \qquad (2.2) \\
&= \prod_{j=1}^{s} P\left(S_{jt} \middle| S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}\right) \qquad (2.3) \\
&= \prod_{j=1}^{s} P\left(S_{jt} \middle| S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right) \qquad (2.4)
\end{aligned}
$$

where (2.2) and (2.3) follow from Assumptions 1.1 and 1.3, respectively. In Eq. (2.4) we decomposed $\boldsymbol{\eta}$ as $\boldsymbol{\eta} = \left(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_s\right)$, where $\boldsymbol{\eta}_j$ are the parameters associated with the conditional distribution $S_{jt} \big| S_{1t-1}, \ldots, S_{st-1}$. As previously with $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$, the vector parameters $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_s$ are variation free, as will become clear later on. Rearranging all terms one has

$$
\begin{aligned}
f\left(y_0, y_1, \ldots, y_n; S_{j0}, S_{j1}, \ldots, S_{jn}; \boldsymbol{\theta}\right) &= \prod_{t=1}^{n} f\left(y_t | \mathbf{S}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}\right) \prod_{t=1}^{n} \prod_{j=1}^{s} P\left(S_{jt} \middle| S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right) \\
&= \prod_{t=1}^{n} f\left(y_t | \mathbf{S}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}\right) \prod_{t=1}^{n} P\left(S_{1t} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right) \\
&\quad \ldots \prod_{t=1}^{n} P\left(S_{st} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right) \qquad (2.5)
\end{aligned}
$$

and the log likelihood is

$$
\begin{aligned}
\log f\left(y_0, y_1, \ldots, y_n; S_{j0}, S_{j1}, \ldots, S_{jn}; \boldsymbol{\theta}\right) &= \sum_{t=1}^{n} \log f\left(y_t | \mathbf{S}_t, \mathcal{F}_{t-1}; \boldsymbol{\psi}\right) \\
&\quad + \sum_{t=1}^{n} \log P\left(S_{1t} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_1\right) \\
&\quad + \cdots + \sum_{t=1}^{n} \log P\left(S_{st} | S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_s\right). \qquad (2.6)
\end{aligned}
$$

This decomposition shows that the parameters can be estimated separately, by maximizing the various expressions in the previous equation, without any loss of consistency or efficiency. Consequently, $\boldsymbol{\psi} = (\boldsymbol{\beta}, \boldsymbol{\delta})$ is estimated, for example, using the ML in Eq. (1.2), and $\boldsymbol{\eta}_j$ $(j = 1, \ldots, s)$ are estimated taking each conditional distribution $S_{jt} \big| S_{1t-1}, \ldots, S_{st-1}$ one at a time, as we will describe in the next section (see for example Eq. (3.2)).

3. Multivariate Markov Chain estimation

The purpose of this section is to describe a method to estimate the parameters $\boldsymbol{\eta}_j$ defined in the log-likelihood expression (2.6). As proved in the previous section, the expression $\sum_{t=1}^{n} \log P\left(S_{jt} \middle| S_{1t-1}, \ldots, S_{st-1}; \boldsymbol{\eta}_j\right)$ can be maximized independently of the other terms contained in the log-likelihood function (2.6). Let $P_j\left(i_0 | i_1, \ldots, i_s\right) \equiv P\left(S_{jt} = i_0 \middle| S_{1,t-1} = i_1, \ldots, S_{s,t-1} = i_s\right)$ where $j \in \{1, 2, \ldots, s\}$ and $i_1, \ldots, i_s \in \{1, 2, \ldots, m\}$. It is well known that modeling these probabilities when s and m are relatively large and the sample size is small or even moderate, is unfeasible because the total number of parameters is $m^s(m-1)$, as can be shown. To overcome this problem Raftery (1985) considered a simplifying hypothesis for modeling high-order Markov chains (HOMC). Recently, Nicolau (in press) proposed an alternative specification, called the MTD-Probit model:

$$P_j\left(i_0\middle|\, i_1,\ldots,i_s\right)=P_j^{\Phi}\left(i_0\middle|\, i_1,\ldots,i_s\right)\equiv\frac{\Phi\left(\eta_{j0}+\eta_{j1}P_{j1}\left(i_0\middle|\, i_1\right)+\cdots+\eta_{js}P_{js}\left(i_0\middle|\, i_s\right)\right)}{\sum_{k=1}^{m}\Phi\left(\eta_{j0}+\eta_{j1}P_{j1}\left(k\middle|\, i_1\right)+\cdots+\eta_{js}P_{js}\left(k\middle|\, i_s\right)\right)}\tag{3.1}$$

where $\eta_{ji}\in\mathbb{R}\,(j=1,\ldots,s;\,i=1,\ldots,m)$ and Φ is the (cumulative) standard normal distribution function. When S_{jt} is the dependent variable the likelihood is

$$\log L=\sum_{i_1i_2\ldots i_s i_0}n_{i_1i_2\ldots i_s i_0}\log\left(P_j^{\Phi}\left(i_0\middle|\, i_1,\ldots,i_s\right)\right)\tag{3.2}$$

and the maximum likelihood estimator is defined, as usual, as $\hat{\eta}_j=\arg\max_{\eta_{j1},\ldots,\eta_{js}}\log L$. The parameters $P_{jk}\left(i_0\middle|\, i_1\right),k=1,\ldots,s$ can be estimated in advance, through the consistent estimators $\hat{P}_{jk}\left(i_0\middle|\, i_1\right)=\frac{n_{i_1i_0}}{\sum_{i_0=1}^{n}n_{i_1i_0}}$ where $n_{i_1i_0}$ is the number of transitions from $S_{k,t-1}=i_1$ to $S_{jt}=i_0$. This procedure greatly simplifies the estimation procedure and does not alter the consistency of the MLE $\hat{\eta}_j$ estimator, as $\hat{P}_{jk}$ is a consistent estimator of P_{jk}.

4. Multi-step forecast model

The previous section described how the probabilities $P\left(S_{jt}=i_0\middle|\, S_{1,t-1}=i_1,\ldots,S_{s,t-1}=i_s\right)$ can be estimated. In this section we introduce the h-step-ahead MMC forecast problem, i.e. $P\left(S_{j,t+h}=k\middle|\, S_{1t}=i_1,\ldots,S_{st}=i_s\right)$. Since we have a homogeneous MMC, the one-step-ahead forecast expression is straightforward, given Assumption 1.2: $P\left(S_{jt+1}=k|S_{1t},\ldots,S_{st}\right)=P\left(S_{jt}=k|S_{1t-1},\ldots,S_{st-1}\right)$.

To obtain the h-step-ahead MMC forecast, we consider two procedures. In the first we start to deduce a general formula for the h-step-ahead MMC forecast that can be recursively computed from the previous forecast. Using the discrete version of Chapman–Kolmogorov equations, the formula of total probability, and Assumptions 1.1–1.3, we have

$$\begin{aligned}
&P\left(S_{jt+h}=k|S_{1t},\ldots,S_{st}\right)\\
&=\sum_{i_1}^{m}\sum_{i_2}^{m}\cdots\sum_{i_s}^{m}P\left(S_{jt+h}=k\middle|\, S_{1t+h-1}=i_1,\ldots,S_{st+h-1}=i_s,S_{1t},\ldots,S_{st}\right)\\
&=\sum_{i_1}^{m}\sum_{i_2}^{m}\cdots\sum_{i_s}^{m}P\left(S_{jt+h}=k\middle|\, S_{1t+h-1}=i_1,\ldots,S_{st+h-1}=i_s\right)\\
&\quad\times\underbrace{P\left(S_{1t+h-1}=i_1|S_{1t},\ldots,S_{st}\right)}_{\text{from }h-1}\underbrace{P\left(S_{2t+h-1}=i_1|S_{1t},\ldots,S_{st}\right)}_{\text{from }h-1}\\
&\quad\times\cdots\times\underbrace{P\left(S_{st+h-1}=i_1|S_{1t},\ldots,S_{st}\right)}_{\text{from }h-1}.
\end{aligned}\tag{4.1}$$

This formula is calculated recursively (notice that it depends on $P\left(S_{jt+h-1}=i_1|S_{1t},\ldots,S_{st}\right),j=1,2,\ldots,s.$). The second procedure is based on the assumption that

$$\begin{aligned}
&P\left(S_{jt+h}=i_0\middle|\, S_{1,t}=i_1,\ldots,S_{s,t}=i_s\right)\\
&=\frac{\Phi\left(\eta_{j0}+\eta_{j1}P\left(S_{jt+h}=i_0\middle|\, S_{1t}=i_1\right)+\cdots+\eta_{js}P\left(S_{jt+h}=i_0\middle|\, S_{st}=i_s\right)\right)}{\sum_{k=1}^{m}\Phi\left(\eta_{j0}+\eta_{j1}P\left(S_{jt+h}=i_0\middle|\, S_{1t}=i_1\right)+\cdots+\eta_{js}P\left(S_{jt+h}=i_0\middle|\, S_{st}=i_s\right)\right)}
\end{aligned}\tag{4.2}$$

which is clearly a natural extension of Eq. (1.4). This expression requires that $P\left(S_{jt+h}=i_0\middle|\, S_{kt}=i_k\right)$ be computed in advance. From the Chapman–Kolmogorov equations and the formula of total probability, it can be easily seen that

$$P\left(S_{jt+h}=i_0\middle|\, S_{kt}=i_k\right)=\sum_{\alpha=1}^{m}P\left(S_{jt+h}=i_0\middle|\, S_{k,t+h-1}=\alpha\right)P\left(S_{k,t+h-1}=\alpha\middle|\, S_{kt}=i_k\right).\tag{4.3}$$

This expression is equal to the element (i_0,i_k) of the matrix product $P^{(jk)}\left(P^{(kk)}\right)^{h-1}$ where $P^{(jk)}$ is a matrix with elements $P\left(S_{jt}=i_0\middle|\, S_{kt-1}=i_k\right)$. We found formula (4.2) computationally easier to implement than (4.1).

We may now establish the algorithm behind the forecast of y_{t+h}:

1. Run the regression model $y_t=\boldsymbol{x}_t'\boldsymbol{\beta}+\boldsymbol{z}_t'\boldsymbol{\delta}+u_t$ and estimate $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ using the OLS method or any other method.
2. Obtain $\hat{\eta}_j=\arg\max_{\eta_{j1},\ldots,\eta_{js}}\log L$ where the log-likelihood refers to Eq. (3.2).

3. From the estimates $\hat{\eta}_j$ calculate $P\left(S_{jt+1} = k|S_{1t}, \ldots, S_{st}\right)$, and derive the expressions $P\left(S_{jt+h} = k|S_{1t}, \ldots, S_{st}\right)$, either recursively from formula (4.1) or from formula (4.2).
4. Finally, obtain the forecast y_{t+h} by calculating $E\left(y_{t+h}|\mathcal{F}_t\right) = E\left(\boldsymbol{x}'_{t+h}\middle|\mathcal{F}_t\right)\boldsymbol{\beta} + E\left(\boldsymbol{z}'_{t+h}\middle|\mathcal{F}_t\right)\boldsymbol{\delta}$.

5. Monte Carlo simulation study

5.1. Monte Carlo simulation study: procedure and design

In this section we evaluate the MMC predictive potential through a Monte Carlo simulation problem. The goal is to construct a model where the MMC, transformed into $s \times (m-1)$ dummy variables (one dummy for each state minus one, for each category), play the role of covariates, seeking to gauge how they help forecast a certain dependent variable.

We consider here a simple process with two categories ($s = 2$) with each one taking values of 1, 2 or 3 ($m = 3$). We simulate the MMC in accordance with the following algorithm:

1. Initialize the process $\{(S_{1t}, S_{2t})\}$ by assigning arbitrary values for S_{10} and for S_{20}.
2. Define two $m^s \times m$ TPMs whose elements are, respectively, the following probabilities

$$\begin{aligned} &P\left(S_{1t} = i_o|\, S_{1t-1} = i_1, S_{2t-1} = i_2\right) \\ &P\left(S_{2t} = i_o|\, S_{1t-1} = i_1, S_{2t-1} = i_2\right) \end{aligned} \tag{5.1}$$

   (see the definition of the data-generation process below).
3. Given the initial values S_{10} and S_{20} (step 1), simulate the multivariate process $\{(S_{1t}, S_{2t})\}$, $t = 1, \ldots, T$ as follows:
   (a) simulate U_1, uniformly distributed on [0, 1];
   (b) let us define $p_i^{[1]} \equiv P\left(S_{1t} = i|\, S_{1t-1} = i_1, S_{2t-1} = i_2\right)$;
   (c) assign a value to S_{1t} according to the rule:

$$S_{1t} = \begin{cases} 1 & \text{if} \quad 0 \le U_1 < p_1^{[1]} \\ 2 & \text{if} \quad p_1 \le U_1 < p_1^{[1]} + p_2^{[1]} \\ 3 & \text{if} \quad p_1 + p_2 \le U_1 < 1 \end{cases}$$

   (d) repeat this procedure for S_{2t} (using $U_2 \sim U(0, 1)$, independent of U_1).
4. Repeat the steps 1–4 until $t = T$.

Thus, we construct our 4 dummy variables, as in (1.1), such that: $z_{jk,t} = \mathcal{I}_{\{S_{jt}=k\}}$, $k = 1, \ldots, m-1$.

We consider the following linear data-generation process (DGP) where

- $\boldsymbol{z}'_t \equiv \begin{bmatrix} z_{11} & z_{12} & z_{21} & z_{22} \end{bmatrix}$, $\boldsymbol{\delta} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}'$, for simplicity,
- $\boldsymbol{x}'_t = \begin{bmatrix} 1 & x_t \end{bmatrix}$ and x_t (such as u_t) is i.i.d. $N(0, 1)$, $\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 \end{bmatrix}'$.

To fully define the DGP, we arbitrarily construct the TMP as follows:

S_{1t-1}		1	1	1	2	2	2	3	3	3
S_{2t-1}		1	2	3	1	2	3	1	2	3
	1	0.1	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.1
S_{1t}	2	0.1	0.4	0.1	0.4	0.8	0.1	0.1	0.4	0.1
	3	0.8	0.1	0.8	0.1	0.1	0.4	0.8	0.1	0.8

For convenience, we assume that S_{1t} and S_{2t} have the same transition probabilities, i.e. that $P(S_{2t} = i_0|\, S_{1t-1} = i_1, S_{2t-1} = i_2) = P(S_{1t} = i_0|\, S_{1t-1} = i_1, S_{2t-1} = i_2)$.

We aim to compare the dependent variable h-step-ahead forecast errors produced by four different hypotheses:

Case 1. The values of dummy variables at $t + h$ are known,

$$\hat{z}^{(1)}_{jkt+h} = z_{jkt+h} \tag{5.2}$$

Case 2. The values of dummy variables at $t + h$ are predicted using the following proposed methodology

$$\hat{z}^{(2)}_{jkt+h} = \hat{P}\left(S_{jt+h} = k\middle|\, S_{1t} = i_1, S_{2t} = i_2\right) \tag{5.3}$$

where $\hat{P}\left(S_{jt+h} = k\middle|\, S_{1t} = i_1, S_{2t} = i_2\right)$ is obtained according to expression (4.3).

Case 3. The values of dummy variables at $t + h$ are predicted using marginal means

$$\hat{z}^{(3)}_{jkt+h} = T^{-1}\sum_{t=1}^{T} z_{jkt}. \tag{5.4}$$

Note that we estimate the event $\mathcal{I}_{\{S_{jt}=k\}}$ using a consistent estimator for the marginal probability, $P\left(S_{jt+h} = k\right)$.

Case 4. The dummies are omitted, i.e $\hat{z}^{(4)}_{jkt+h} \equiv 0$.

Out-of-sample forecasts were generated by the so-called recursive (expanding windows) forecasts. An initial sample using data from $t = 1$ to $T = 1000$ is used to estimate the models, and h-step ahead out-of-sample forecasts are produced starting

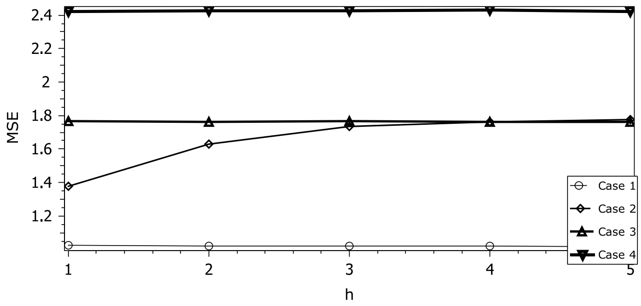

Fig. 1. Results of the forecast errors MSE_{lh}.

at time $T = 1000$. The sample is increased by one, the models are re-estimated, and h-step ahead forecasts are produced starting at $T + 1$. This procedure is repeated 1000 times, i.e. we considered 1000 out-of-sample forecasts, and the forecasting time horizon was defined as $h = 1, 2, 3, 4, 5$. Lastly, we assessed the quality of the forecast using the statistics $MSE_{lh} = N^{-1}\sum_{t=T}^{T+999}\hat{e}^2_{l,t+h}$ where $N = 1000$ is the number of replicas considered in the experiment and $e_{l,t+h}$ is the forecast error produced by model l ($l = 1, 2, 3, 4$) at the hth forecast step, i.e. $e_{l,t+h} \equiv y_{t+h} - \hat{y}^{(l)}_{t+h}$, where $\hat{y}^{(l)}_{t+h} \equiv \boldsymbol{x}'_{t+h}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{z}}^{(l)'}_{t+h}\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{z}}^{(l)'}_{t+h} \equiv \begin{bmatrix} z^{(l)}_{11t+h} & z^{(l)}_{12t+h} & z^{(l)}_{21t+h} & z^{(l)}_{22t+h} \end{bmatrix}$, for $l = 1, 2, 3, 4$.

5.2. Monte Carlo simulation study: discussion of results

In this section we report the results of the Monte Carlo study presented in the previous section, investigating the potential forecast gains of a dependent variable, derived by processing categorical interrelated regressors as an MMC, i.e. by exploiting intra and inter-transition probabilities between categorical regressors. Fig. 1 presents the MSE_{lh} for $l = 1, 2, 3, 4$ and $h = 1, 2, 3, 4$.

As expected, case 1 presents the best results, since the forecast of y_{t+h} is based on the actual values z_{jkt+h}, and case 4 the worst result, since the dummies z_{jkt+h} were simply ignored. Case 2 uses the proposed methodology, and hence explores the intra and inter-transition probabilities between categorical regressors; it clearly produces better results than case 3, in which the forecasts are based on the estimate of the marginal probabilities $P\left(S_{jt+h} = k\right)$. To confirm the advantage of the proposed method over the marginal probabilities we carried out the Diebold and Mariano (2002) (DM) test, that allows us to assess the significance of the MSE difference between those models. As is known, the DM can be trivially calculated by regression $\hat{e}^2_{3,t+h} - \hat{e}^2_{2,t+h}$ on an intercept, using heteroskedasticity and autocorrelation robust (HAC) standard errors. Our results (available upon request) show that the proposed model outperforms the forecasts based on the marginal mean for $h = 1$ and $h = 2$ (p-value zero), and possibly $h = 3$ (p-value 0.08). When h increases, the advantage of using our model dissipates, which is to be expected taking into account that in stationarity and weak dependence assumptions, the conditional probabilities converge into the stationary probabilities, i.e. $P\left(S_{jt+h} = i_0 \middle| S_{1t} = i_1, S_{2t} = i_2\right) \to P\left(S_{jt} = i_0\right)$ as $h \to \infty$.

6. Conclusions

This paper proposed a new concept by using MMC as covariates in a regression model in order to improve the forecast error of a certain dependent variable, provided it is caused by the MMC. Traditionally, the published literature only addresses the MMC as an end in itself. In the context of an endogenous variable that depends on some time-dependent categorical or discrete variables, we show that taking advantage of the information about the past state interactions between the categorical variables through an MMC specification via modeling the intra and inter-transition probabilities within and between data categories, may lead to a substantial forecasting improvement of that endogenous variable, as the Monte Carlo experiment has shown.

Acknowledgment

This research was supported by the Fundação para a Ciência e a Tecnologia (Science and Technology Foundation) Project PEstOE/EGE/UI0491/2013.

References

Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. J. Bus. Econom. Statist. 20 (1).
Engle, R.F., Hendry, D.F., Richard, J.-F., 1983. Exogeneity. Econometrica 277–304.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 357–384.
Nicolau, J., 2014. A method to estimate multivariate markov chains. Scand. J. Stat. http://dx.doi.org/10.1111/sjos.12087. (in press).
Raftery, A.E., 1985. A model for high-order markov chains. J. R. Stat. Soc. Ser. B Stat. Methodol. 528–539.