Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 134 (2006) 553–577
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Markov-switching model selection using Kullback–Leibler divergence

Aaron Smith[a,*], Prasad A. Naik[b], Chih-Ling Tsai[b,c]

[a]*Department of Agricultural and Resource Economics, University of California, One Shields Avenue, Davis, CA 95616, USA*
[b]*Graduate School of Management, University of California, Davis, USA*
[c]*Guanghua School of Management, Peking University, PR China*

Available online 29 August 2005

Abstract

In Markov-switching regression models, we use Kullback–Leibler (KL) divergence between the true and candidate models to select the number of states and variables simultaneously. Specifically, we derive a new information criterion, Markov switching criterion (MSC), which is an estimate of KL divergence. MSC imposes an appropriate penalty to mitigate the over-retention of states in the Markov chain, and it performs well in Monte Carlo studies with single and multiple states, small and large samples, and low and high noise. We illustrate the usefulness of MSC via applications to the U.S. business cycle and to media advertising.

JEL classification: C22; C52.

Keywords: Advertising effectiveness; Business cycles; EM algorithm; Hidden Markov models; Information criterion; Markov-switching regression.

1. Introduction

Economic systems often experience shocks that shift them from their present state into another state; for example, nations lurch into recession, government regimes

*Corresponding author. Tel.: 530 752 2138; fax: 530 752 5614.
E-mail address: adsmith@ucdavis.edu (A. Smith).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.07.005

change over time, and financial markets exhibit booms and crashes. These states tend to be stochastic and dynamic: if they occur once, they probably recur. To capture such probabilistic state transitions over time, Markov-switching models provide an analytical framework. In economics, Markov-switching models have been used for investigating the US business cycle (Hamilton, 1989), foreign exchange rates (Engel and Hamilton, 1990), stock market volatility (Hamilton and Susmel, 1994), real interest rates (Garcia and Perron, 1996), corporate dividends (Timmermann, 2001), the term structure of interest rates (Ang and Bekaert, 2002a), and portfolio allocation (Ang and Bekaert, 2002b), among others. Outside of economics, Markov-switching models find application in diverse fields such as computational biology (e.g., Durbin et al., 1998 for gene sequencing), computer vision (Bunke and Caelli, 2001), and speech recognition (Rabiner and Juang, 1993).

To estimate Markov-switching models, Baum and his colleagues (Baum and Petrie, 1966; Baum et al., 1970) developed the forward–backward algorithm, which was extended to encompass general latent variable models under the expectation–-maximization (EM) principle (see Dempster et al., 1977). If the number of states in Markov-switching models is known, the EM algorithm yields consistent parameter estimates, and statistical inference proceeds via standard maximum-likelihood theory (e.g., Bickel et al., 1998). If the number of states is not known, however, the likelihood ratio test to infer the true number of states breaks down because regularity conditions do not hold (see Hartigan, 1977; Hansen, 1992; Garcia, 1998).

The number of states is often not known a priori, so we propose applying KL divergence to determine it. We note that KL divergence has been used in various model selection contexts (see, e.g., Sawa, 1978; Leroux, 1992; Sin and White, 1996; Burnham and Anderson, 2002). Specifically, Akaike's information criterion (AIC, see Akaike, 1973) provides an estimate of KL distance but, in Markov-switching models, it misleads the users into selecting too many states (see Section 4.2). Consequently, one fits spurious regressions in nonexistent states; this misspecification results in incorrect inclusion of variables, which reduces the accuracy of estimated parameters and lowers the precision of model forecasts. Hence, the problem of simultaneous determination of the number of states to retain in the Markov chain and the variables to include in the regression model for each retained state remains open.

The objective of this paper is to develop a new information criterion for simultaneous selection of states and variables in Markov switching models. To accomplish this goal, we obtain an explicit approximation to the KL distance for the class of Markov switching regression models. The resulting Markov switching criterion (MSC) imposes an appropriate penalty, and so it mitigates the over-retention of states in the Markov chain and alleviates the tendency to over-fit the number of variables in each state. Moreover, in Monte Carlo studies, MSC performs well in single and multiple states, small and large samples, and low and high noise. Finally, it not only applies to Markov-switching regression models, but also performs well in Markov-switching autoregression models.

We present two empirical applications of MSC to understand (a) the business cycle in the US economy and (b) the effectiveness of media advertising. In the