



1-2001

Autocovariance Structure of Markov Regime Switching Models and Model Selection

Jing Zhang
University of Pennsylvania

Robert A. Stine
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/statistics_papers

 Part of the [Statistics and Probability Commons](#)

Recommended Citation

Zhang, J., & Stine, R. A. (2001). Autocovariance Structure of Markov Regime Switching Models and Model Selection. *Journal of Time Series Analysis*, 22 (1), 107-124. <http://dx.doi.org/10.1111/1467-9892.00214>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/statistics_papers/437
For more information, please contact repository@pobox.upenn.edu.

Autocovariance Structure of Markov Regime Switching Models and Model Selection

Abstract

We show that the covariance function of a second-order stationary vector Markov regime switching time series has a vector ARMA(p, q) representation, where upper bounds for p and q are elementary functions of the number of regimes. These bounds apply to vector Markov regime switching processes with both mean–variance and autoregressive switching. This result yields an easily computed method for setting a lower bound on the number of underlying Markov regimes from an estimated autocovariance function.

Keywords

ARMA, AIC, BIC

Disciplines

Statistics and Probability

Autocovariance Structure of Markov Regime Switching Models and Model Selection

Jing Zhang and Robert A. Stine

Department of Statistics

The Wharton School of the University of Pennsylvania

Philadelphia, PA 19104-6302

September 1, 1999

Abstract

We show that the covariance function of a second-order stationary vector Markov regime switching time series has a vector $ARMA(p, q)$ representation, where upper bounds for p and q are elementary functions of the number of regimes. These bounds apply to vector Markov regime switching processes with both mean-variance and autoregressive switching. This result yields an easily computed method for setting a lower bound on the number of underlying Markov regimes from an estimated autocovariance function.

Key words: ARMA, AIC, BIC.

1 Introduction

We show that the autocovariance structure of a model belonging to a general class of second order stationary Markov regime switching processes is that of a vector autoregressive moving average (*VARMA*) whose orders p and q are bounded above by elementary functions of the number of Markov regimes k (see Theorems 3 and 4 in Section 3). This result applies to models with both mean-variance switching as well as switching among autoregressive regimes, unifying and extending previous work. In the case of a mean-variance switching process, the orders $p, q < k$. For models switching among autoregressions, the bounds are elementary functions of the dimension of the process, the number of regimes, and the maximum order of autoregression. As the sample autocovariances are more easily calculated than maximum likelihood estimates of the model parameters, these bounds can be very valuable in model selection. In particular, our result yields an estimate of the lower bound on the number of regimes. Such a lower bound is particularly relevant in light of the result of Donoho (1988) which discusses the inability to build two-sided confidence intervals for the complexity of certain models.

Given a model structure with a known number of regimes, efficient and sophisticated estimation and forecasting schemes have been successfully developed using Markov regime switching time series models (also known as hidden Markov models). These have been applied in a variety of fields including speech recognition (Juang and Rabiner 1990), DNA composition (Churchill 1989), ion channels (Chung *et al.* 1990, Fredkin and Rice 1992), analysis of business cycles (Hamilton 1989, 1990) and modeling stock market and asset returns

(Turner, Startz, and Nelson 1990). However, the techniques in the literature for choosing the number of regimes k are incomplete and often difficult to apply. The conventional likelihood ratio test fails if one tries to fit a k -regime model when the true process has $k - 1$ regimes since under the null hypothesis the parameters that describe the k -th regime are unidentified. Hansen (1992, 1996) proposed a test that avoids this problem, but his test only bounds the likelihood ratio and requires three-dimensional grid search. Hamilton (1996) used another approach that treats the $(k - 1)$ -regime model as the null, and he conducted a variety of tests of whether a k -regime model is needed. Leroux and Puterman (1992) and Rydén (1995) studied the use of traditional criteria such as AIC or BIC for determining the number of regimes in mean-variance switching models, and the latter proved that AIC and BIC will not underestimate the number of the regimes.

A simpler approach to model selection exploits the relationship of the covariances of these models to those of *ARMA* models. Several authors have studied the autocovariance structure for the scalar case. Karlsen (1990) shows that the autocovariances of a second order stationary scalar process which switches among k *AR*(1) regimes have an *ARMA*(p, q) representation, where $p \leq k$ and $q \leq k - 1$. He conjectures that this is also true for vector case. Anděl (1993) proves that a two-state scalar Markov mean switching model has an *ARMA*(1, 1) representation. Recently, for the k -state scalar Markov mean switching model, Poskitt and Chung (1996) show that if the Markov chain has a non-singular transition matrix P , the autocovariance function is that of an *ARMA*($k - 1, k - 1$) process for most regimes.

The rest of this article is organized as follows: In Section 2, we state the specifics of the

models to be considered and give a characterization of a $VARMA(p, q)$ process in terms of autocovariances. For completeness, we also quote conditions for second-order stationarity of $\{Y_t\}$ obtained by Karlsen (1990). In Section 3, we show that $\{Y_t\}$ has a $VARMA(p, q)$ representation where the upper bounds for p and q are elementary functions of the number of the Markov regimes. In Section 4, we discuss the implications of our result for model selection. Section 4 includes simulations which compare several methods for identifying the number of states of the Markov chain. In Section 5, we illustrate our results on the exchange rate data from Engel and Hamilton (1990). We close in Section 6 with a brief discussion.

2 Preliminaries

We consider discrete-time Markov regime switching models with the following vector autoregressive form (Karlsen 1990):

$$Y_t = U_{s_t} + A_{s_t}(Y_{t-1} - U_{s_{t-1}}) + \Sigma_{s_t} V_t, \quad t \in \mathbf{Z}, \quad (1)$$

where \mathbf{Z} denotes the integers, Y_t is an m -dimensional random vector, and the parameter process $\{H_t\} = \{(A_{s_t}, \Sigma_{s_t}, U_{s_t})\}$ is governed by a k -state irreducible stationary ergodic Markov chain $\{s_t\}$; when $s_t = i$, $H_t = (A_i, \Sigma_i, U_i)$. Throughout this paper, we assume $\{s_t\}$ has transition matrix P with elements $p_{ij} = P(s_{t+1} = j \mid s_t = i)$ and stationary distribution $\Pi = (\pi_1, \dots, \pi_k)'$. We also assume that $\{(s_t, V_t)\}$ is a strictly stationary process defined on some probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and $\{s_t\}$ is independent of $\{V_t\}$. In addition, we assume

that $\{V_t\}$ have mean zero and are uncorrelated,

$$\begin{cases} E(V_t) & = & 0 \\ E(V_{t_1} V_{t_2}^T) & = & \delta_{t_1, t_2} I_m, \end{cases} \quad (2)$$

where I_m is the $m \times m$ identity matrix and δ is the Kronecker function, i.e., $\delta_{t_1, t_2} = 1$ if $t_1 = t_2$ and 0 otherwise. Note that assuming a first-order vector autoregression is not restrictive since any (scalar or vector) autoregression can be written as a $VAR(1)$ by stacking. This formulation includes the scalar mean-switching models ($A_{s_t} = 0$) of Anděl (1993) and Poskitt and Chung (1996), and the autoregressive switching models of Hamilton (1996).

We next offer a characterization of a $VARMA(p, q)$ process in terms of its autocovariance function. We then review Karlsen's (1990) condition for second-order stationarity. Recall the definition of a second-order stationary m -dimensional $VARMA(p, q)$ process $\{Y_t\}$:

Definition 1 *A second-order stationary m -dimensional process $\{Y_t\}$ is a stable and invertible $VARMA(p, q)$ process if it satisfies the following:*

$$\Phi(L)Y_t = \Theta(L)V_t, \quad (3)$$

where L is the lag operator; $\Phi(L) = I - A_1L - \dots - A_pL^p$, $A_p \neq 0$; $\Theta(L) = I - M_1L - \dots - M_qL^q$, $M_q \neq 0$; V_t is white noise with $E(V_t) = 0$, $E(V_t V_t^T) = G$, $\det G \neq 0$, $E(V_t V_s^T) = 0$ if $t \neq s$. To avoid redundancy, $\Phi(L)$ and $\Theta(L)$ are coprime; that is, if $\Phi(L) = U(L)\phi(L)$, $\Theta(L) = U(L)\theta(L)$, and $U(L)$ is not a constant matrix, then $U(L)$ is a unimodular matrix with $|U(L)| = 1$. Furthermore, the polynomials $\det(z^p I_m - z^{p-1}A_1 - \dots - A_p)$ and $\det(z^q I_m - z^{q-1}M_1 - \dots - M_q)$ have all of their zeros inside the unit circle.

This definition implies that the orders (p, q) are minimal: no smaller (p, q) and matrices $A_i, M_j, i = 1, 2, \dots, p, j = 1, 2, \dots, q$ exist such that (3) holds. The above representation of a $VARMA(p, q)$ may not be unique. Lütkepohl (1993) discusses conditions for identifiability. Here we are more concerned with the minimal order (p, q) than the uniqueness of A_i, M_j .

The following theorem characterizes the minimal $VARMA(p, q)$ representation in terms of its autocovariance function. This theorem extends the scalar result of Beguin, Gourierous, and Monfort (1980) to vectors.

Theorem 1 *A zero mean, second-order stationary m -dimensional process $\{Y_t\}$ has a minimal $VARMA(p, q)$ representation if and only if the covariances $\gamma(h) = E(Y_t Y_{t-h}^T), h \in \mathbf{Z}$, satisfy a difference equation of minimal order p with minimal rank $q + 1$. That is, (p, q) are the smallest pair such that there exist $m \times m$ matrices $A_1, A_2, \dots, A_p, A_p \neq 0$ such that:*

$$\gamma(h) - A_1 \gamma(h-1) - \dots - A_p \gamma(h-p) \begin{cases} = 0 & \text{for } h \geq q+1 \\ \neq 0 & \text{for } h = q. \end{cases} \quad (4)$$

Proof: The proof parallels that of Beguin *et al.* (1980). See the Appendix for details. ■

As an illustration of the techniques we will use in the next section, we prove the following lemma using the above theorem.

Lemma 1 *If the autocovariance function $\gamma(h)$ of an m -dimensional second-order stationary process $\{Y_t\}$ can be written as*

$$\gamma(h) = a^T Q^h b \quad \text{for } h \geq 0, \quad (5)$$

where a, b are $n \times m$ nonzero matrices and Q is an $n \times n$ matrix, then $\{Y_t\}$ has a $VARMA(p, q)$ representation where $p \leq n$ and $q \leq n - 1$.

Proof: The Cayley-Hamilton theorem implies there exist $\phi_1, \phi_2, \dots, \phi_n$ such that:

$$Q^n - \phi_1 Q^{n-1} - \dots - \phi_n I_n = f(Q) = 0, \quad (6)$$

where $f(\lambda) \stackrel{\text{def}}{=} \lambda^n - \phi_1 \lambda^{n-1} - \dots - \phi_n$ is the characteristic polynomial of Q . The above equation implies

$$a^T (Q^n - \phi_1 Q^{n-1} - \dots - \phi_n I_n) b = 0, \quad (7)$$

and thus

$$, (n) - \phi_1, (n-1) - \dots - \phi_n, (0) = 0. \quad (8)$$

The result then follows from Theorem 1. ■

Lemma 1 gives an upper bound on the orders of the autoregression and moving average terms. The exact order of the $VARMA(p, q)$ representation depends on the configuration of Q, a , and b . For example in the scalar case:

1. When Q is singular, the process is

(a) White noise when either $a^T Q = 0$ or $Qb = 0$.

(b) $MA(q)$, $q \leq n - 1$, when there exists a q such that $0 < q < n - 1$ and $q = \min\{k_1, k_2, k_3 \in \mathbf{Z} : Q^{k_1} = 0, a^T Q^{k_2} = 0, Q^{k_3} b = 0\}$.

(c) $ARMA(p, q)$ with $p, q \leq n - 1$. Let $f(\lambda)$ be the minimal polynomial of Q , where the minimal polynomial $f(\lambda)$ of an $n \times n$ matrix A is defined as the polynomial with

the smallest degree of those satisfying $g(A) = 0$. Assume $f(\lambda) = \lambda^m - \phi_1 \lambda^{m-1} - \dots - \phi_r \lambda^{m-r}$. It is well known (*e.g.*, Marcus and Minc 1964) that $m - r > 0$ when A is singular. Combining $f(Q) = 0$ and Lemma 1 implies that $\{y_t\}$ has an $ARMA(p, q)$ representation where $p \leq r < m \leq n$ and $q + 1 \leq m < n$.

2. When Q is nonsingular, the process is $ARMA(p, q)$, $p \leq n, q \leq n - 1$. Let $f(\lambda)$ be the minimal polynomial of Q and m denote the degree of $f(\lambda)$. By Lemma 1, we have $p \leq m \leq n$ and $q + 1 \leq m \leq n$.

Next, for completeness, we review Karlsen's (1990) conditions for the second-order stationarity of $\{Y_t\}$ in the regime-switching model (1) and his expression for calculating the autocovariances of $\{Y_t\}$. We need the following matrices:

1. $D_0 \stackrel{\text{def}}{=} \text{diag}\{(\Sigma_j \otimes \Sigma_j), j = 1, 2, \dots, k\}$: a diagonal block matrix of total order $km^2 \times km^2$ where each block $\Sigma_j \otimes \Sigma_j$ has dimension $m^2 \times m^2$.
2. $D_1 \stackrel{\text{def}}{=} \text{diag}\{(I_m \otimes A_j), j = 1, 2, \dots, k\}$: a diagonal block matrix with the j th diagonal block repeating m copies of A_j .
3. $D_2 \stackrel{\text{def}}{=} \text{diag}\{(A_j \otimes A_j), j = 1, 2, \dots, k\}$.

Let $\{X_t\}$ be the solution of $X_t = A_{s_t} X_{t-1} + \Sigma_{s_t} V_t$. Then the solution for (1) can be written as $Y_t = X_t + U_{s_t}$. The theorem below gives a sufficient condition for second-order stationarity of $\{Y_t\}$:

Theorem 2 (Karlsen 1990) *Let $F_0 = D_0(\Pi \otimes I_{m^2})$ and $F_j = D_j(P^T \otimes I_{m^2})$, $j = 1, 2$. If the spectral radius of F_2 , denoted $\rho(F_2)$, is less than 1 (*i.e.*, all the eigenvalues of F_2 are inside*

the unit disk), then $\{(X_t, Y_t)\}$ is second-order stationary. The covariance structure is given by ($h \geq 0$):

$$\begin{cases} \text{vec } ,_X(h) &= (1_k^T \otimes I_{m^2}) F_1^h (I_{km^2} - F_2)^{-1} F_0 \text{vec } I_m \\ ,_Y(h) &= ,_X(h) + ,_U(h), \end{cases} \quad (9)$$

where 1_k denotes the $k \times 1$ vector of 1's.

For the scalar case, we have the following

Corollary 1 (Karlsen 1990) *Let $m = 1$ and assume that the conditions (2) hold. Let $D_0 = \text{diag}\{\sigma_i^2, i = 1, 2, \dots, k\}$, $D_1 = \text{diag}\{\theta_i, i = 1, 2, \dots, k\}$ and $D_2 = D_1^2$. If all the eigenvalues of PD_1^2 are inside the unit disk, then $\{Y_t\}$ is second-order stationary and*

$$,_Y(h) = a^T (PD_1)^h 1_k, \quad (10)$$

where

$$a = (I - D_1^2 P^T)^{-1} (D_0 \Pi).$$

Example 1 For the scalar case with $k = 2$ (a two-state zero mean switching scalar $AR(1)$ model with $y_t = \theta_{s_t} y_{t-1} + \varepsilon_{s_t}$), we have the following: $D_0 = \text{diag}\{\sigma_1^2, \sigma_2^2\}$, $D_1 = \text{diag}\{\theta_1, \theta_2\}$ and

$$F_k^T = PD_1^k = \begin{pmatrix} p_{11}\theta_1^k & p_{12}\theta_2^k \\ p_{21}\theta_1^k & p_{22}\theta_2^k \end{pmatrix}, \quad k = 1, 2.$$

The characteristic polynomial of F_2 can be calculated as follows

$$\det(\lambda I - F_2) = \lambda^2 - a\lambda + b,$$

where $a = p_{11}\theta_1^2 + p_{22}\theta_2^2$ and $b = \theta_1^2\theta_2^2(p_{11} + p_{22} - 1)$. The condition $\rho(F_2) < 1$ can be formulated in terms of a and b :

$$\rho(F_2) < 1 \iff a - b < 1, a + b > -1 \text{ and } |b| < 1. \quad (11)$$

The above condition is more general than the naive condition:

$$\max\{|\theta_1|, |\theta_2|\} < 1.$$

For example, as long as the probability of staying in state 2 is not large, $|\theta_2|$ can be greater than 1. For example, if

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.5 & 0 \\ 0 & -1.5 \end{pmatrix},$$

then $\rho(F_2) = 0.578 < 1$, yet $\{Y_t\}$ is still stationary even though the coefficient $\theta_2 = -1.5$.

Remark 1 For a constant coefficient $VAR(1)$,

$$Y_t = AY_{t-1} + V_t, \quad (12)$$

the condition for stationarity is that all eigenvalues of the $m \times m$ matrix A lie inside the unit disk, so the condition in Theorem 2 is a natural generalization of the usual $VAR(1)$ case. However, the condition in Theorem 2 is only sufficient. Holst *et al.* (1994) use simulations to exhibit processes which are stationary even though $\rho(F_2) > 1$.

3 Autocovariance of Markov Regime Switching Models

3.1 Mean-Variance Switching Models

Consider the m -dimensional Markov mean-variance switching models

$$Y_t = U_{s_t} + \Sigma_{s_t} V_t, \tag{13}$$

where $\{s_t\}$ is the k -state hidden Markov chain satisfying the assumptions stated in Section 1. The Markov chain $\{s_t\}$ governs the mean and scale processes $\{U_{s_t}\}$ and $\{\Sigma_{s_t}\}$. $\{V_t\}$ is white noise with $E(V_t V_t^T) = I_m$, and independent of $\{s_t\}$.

Let D be the diagonal matrix $diag\{\pi_1, \pi_2, \dots, \pi_k\}$ with the stationary distribution of $\{s_t\}$ on its diagonal, $P_\infty = \lim_{n \rightarrow \infty} P^n$, and $Q = P - P_\infty$. The following are well-known facts:

$$P_\infty = \mathbf{1}_k \Pi^T, \tag{14}$$

$$P_\infty^n = P^n P_\infty = P_\infty P^n = P_\infty, \quad n = 1, 2, \dots \tag{15}$$

$$Q^n = P^n - P_\infty, \tag{16}$$

where (16) can be proved by induction using (14) and (15).

Let $U = (U_1, U_2, \dots, U_k)$ and $W = (\Sigma_1, \Sigma_2, \dots, \Sigma_k)$ denote matrices holding the means and scaling parameters of the k regimes. Since $\{s_t\}$ is second-order stationary, so is $\{Y_t\}$. The following theorem gives a characterization of the autocovariances of $\{Y_t\}$.

Theorem 3 *The process $\{Y_t\}$ defined in (13) has a VARMA(p, q) representation where*

$p \leq k - 1$ and $q \leq k - 1$ and

$$\begin{cases} ,_Y(0) = UD(I - P_\infty)U^T + W(D \otimes I_m)W^T \\ ,_Y(h) = U(Q^T)^hDU^T \quad \text{for } h > 0 \\ ,_Y(h) = [,_Y(-h)]^T \quad \text{for } h < 0. \end{cases} \quad (17)$$

Proof: Note that it is slightly easier to calculate $E(Y_t Y_{t+h}^T) = ,_Y(-h)$. Direct calculations give

$$\begin{aligned} E(U_{s_t}) &= \sum_{i=1}^k U_i P(s_t = i) = U\Pi , \\ E(U_{s_t} U_{s_t}^T) &= \sum_{i=1}^k U_i U_i^T P(s_t = i) = UDU^T , \\ E(U_{s_t})E(U_{s_t}^T) &= U\Pi(U\Pi)^T = UDP_\infty U^T , \\ E(U_{s_t} U_{s_{t+h}}^T) &= \sum_{i=1}^k \sum_{j=1}^k U_i U_j^T P(s_t = i, s_{t+h} = j) = UDP^h U^T . \end{aligned} \quad (18)$$

Thus we have

$$\begin{aligned} ,_{U_{s_t}}(0) &= UD(I - P_\infty)U^T , \\ ,_{U_{s_t}}(-h) &= UD(P - P_\infty)^h U^T = UDQ^h U^T \quad \text{for } h > 0, \end{aligned} \quad (19)$$

and

$$\begin{aligned} ,_Y(-h) &= E(Y_t Y_{t+h}^T) - E(Y_t)E(Y_{t+h})^T \\ &= E(U_{s_t} U_{s_{t+h}}^T) + E(U_{s_t} (\sum_{s_{t+h}} V_{t+h})^T) \\ &\quad + E(\sum_{s_t} V_t U_{s_{t+h}}^T) + E(\sum_{s_t} V_t (\sum_{s_{t+h}} V_{t+h})^T) \\ &\quad - E(U_{s_t})E(U_{s_{t+h}}^T) . \end{aligned} \quad (20)$$

Note that $\{V_t\}$ is white noise and is independent of the Markov chain $\{s_t\}$, and thus

$$\begin{aligned} ,_Y(h) &= [,_Y(-h)]^T = [,_{U_{s_t}}(-h)]^T \quad \text{for } h > 0 , \\ ,_Y(0) &= \text{Var}(Y_t) = \text{Var}(U_{s_t}) + \text{Var}(\sum_{s_t} V_t), \end{aligned}$$

where

$$\begin{aligned}
\text{Var}(\Sigma_{s_t} V_t) &= \text{E}[(\Sigma_{s_t} V_t)(\Sigma_{s_t} V_t)^T] \\
&= \sum_{i=1}^k \Sigma_i \Sigma_i^T P(s_t = i) \\
&= W(D \otimes I_m)W^T.
\end{aligned} \tag{21}$$

The above equations yield (17). For $h > 0$, $\gamma_Y(h) = U(Q^T)^h D U^T$, which is in the form specified in Lemma 1. Let $\lambda_i, i = 1, 2, \dots, k$ be the eigenvalues of P^T , then the eigenvalues of Q^T are $0, \lambda_i - \lim_{n \rightarrow +\infty} \lambda_i^n = \lambda_i, i = 2, 3, \dots, k$ since $\lambda_1 = 0$ and $\|\lambda_i\| < 1$ for $i = 2, \dots, k$ by the assumption of the ergodicity of $\{s_t\}$. Since Q^T is singular, its minimal polynomial can be written as $f(x) = x^k - \phi_1 x^{k-1} - \dots - \phi_{k-1} x$ where ϕ_{k-1} may be 0. An argument similar to that used for Lemma 1 leads to

$$\gamma_Y(k) - \phi_1 \gamma_Y(k-1) - \dots - \phi_{k-1} \gamma_Y(1) = 0, \tag{22}$$

and the result follows from Theorem 1. The proof is slightly different from that for Lemma 1; the minimal polynomial here has a zero constant term since Q is singular. ■

Remark 2 Poskitt and Chung (1996) show that when P is non-singular, the autocovariance function of $\{Y_t\}$ is that of an $ARMA(k-1, k-1)$. However, their proof does not apply to cases where the states have special structure. In such cases, the autocovariance structure is not $ARMA(k-1, k-1)$. To see this, consider the autocovariance $\gamma_Y(h)$ in (17) for a three-state scalar Markov mean switching model. Assume the eigenvalues λ_2, λ_3 of P^T are distinct and non zero, and choose the mean vector $D U^T$ to be the eigenvector of Q^T corresponding to λ_2 . Then $Q^T(Q^T - \lambda_2 I) D U^T = 0$, so that by Theorem 1 $\{Y_t\}$ has an

$ARMA(p, q)$ representation where $p \leq 1, q \leq 1$, not $ARMA(2, 2)$ as claimed by Poskitt and Chung (1996). The exact order (p, q) depends not only on the number of states, but also on the configuration of state vectors and transition matrix. When the mean vector DU^T lies in subspaces of dimension $k - 1$ generated by the eigenvectors of Q^T , reduction of the order (p, q) from $k - 1$ is possible.

3.2 $VAR(p)$ Switching Models

Consider the following m -dimensional Markov regime switching $VAR(p)$:

$$Y_t = A_{s_t}^{(1)}Y_{t-1} + A_{s_t}^{(2)}Y_{t-2} + \cdots + A_{s_t}^{(p)}Y_{t-p} + \Sigma_{s_t}V_t. \quad (23)$$

Write the above equation in vector form as:

$$\mathcal{Y}_t = \mathcal{A}_{s_t}\mathcal{Y}_{t-1} + \mathcal{B}_{s_t}\mathcal{V}_t, \quad (24)$$

where $\mathcal{Y}_t = (Y_t^T, Y_{t-1}^T, \dots, Y_{t-p+1}^T)^T$, $\mathcal{V}_t = (V_t^T, 0, 0, \dots, 0)^T$ and $\mathcal{A}_{s_t}, \mathcal{B}_{s_t}$ are the following $mp \times mp$ matrices

$$\mathcal{A}_{s_t} = \begin{pmatrix} A_{s_t}^{(1)} & A_{s_t}^{(2)} & \cdots & A_{s_t}^{(p-1)} & A_{s_t}^{(p)} \\ I_m & 0 & \cdots & 0 & 0 \\ 0 & I_m & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & I_m & 0 \end{pmatrix}, \quad (25)$$

$$\mathcal{B}_{st} = \begin{pmatrix} \Sigma_{st} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \quad (26)$$

In order to apply Theorem 2 to get the covariance structure of $\{\mathcal{Y}_t\}$, we need $D_i, i = 0, 1, 2$, which are as those in Section 2 with A_i, Σ_i replaced by $\mathcal{A}_i, \mathcal{B}_i$ and m replaced by mp . Now let $F_0 = D_0(\Pi \otimes I_{(mp)^2})$ and $F_j = D_j(P^T \otimes I_{(mp)^2}), j = 1, 2$. By Theorem 2, if the spectral radius $\rho(F_2)$ of F_2 is less than 1, then $\{\mathcal{Y}_t\}$ is second-order stationary with covariance structure given by ($h \geq 0$):

$$vec, \mathcal{Y}(h) = (1_k^T \otimes I_{(mp)^2})F_1^h(I_{k(mp)^2} - F_2)^{-1}F_0 vec I_{mp}. \quad (27)$$

Note $Y_t = C\mathcal{Y}_t$ where C is the $m \times mp$ matrix

$$C = (I_m, 0, \dots, 0). \quad (28)$$

Since $vec(ABC) = (C^T \otimes A)vec B$ for any matrices A, B, C , it is easy to see from Theorem 2 that

$$\begin{aligned} vec, \mathcal{Y}(h) &= vec \{CE(\mathcal{Y}_t\mathcal{Y}_{t-h}^T)C^T\} = vec \{C, \mathcal{Y}(h)C^T\} \\ &= (C \otimes C)vec, \mathcal{Y}(h) \\ &= QF_1^h R, \end{aligned} \quad (29)$$

where

$$Q = (C \otimes C)(1_k^T \otimes I_{(mp)^2}),$$

$$R = (I_{k(mp)^2} - F_2)^{-1} F_0 \{vec I_{mp}\}. \quad (30)$$

Thus by (29) and the characteristic polynomial $f(\lambda) = \lambda^r - \phi_1 \lambda^{(r-1)} - \dots - \phi_r$ of F_1 we get

$$vec_{, Y}(h) - \phi_1 vec_{, Y}(h-1) - \dots - \phi_r vec_{, Y}(h-r) = 0, \text{ for } h-r \geq 0, \quad (31)$$

where $r = k(mp)^2$ is the dimension of F_1 . Equation 31 implies that

$$, Y(h) - A_{1, Y}(h-1) \dots - A_{r, Y}(h-r) = 0, \text{ for } h-r \geq 0, \quad (32)$$

where A_i is the diagonal matrix with ϕ_i on the diagonal. Using Theorem 1, we have

Theorem 4 *The process $\{Y_t\}$ defined in (23) has a VARMA(p', q') representation where $p' \leq k(mp)^2$ and $q' \leq k(mp)^2 - 1$.*

4 Implications for Model Selection

The available methods of model selection and identification for Markov regime-switching time series models are primarily of two types: (1) hypothesis testing based on comparisons of likelihood of the competing models, e.g., Hansen (1992, 1996), Gong and Mariano (1995), and Hamilton (1996); and (2) information criteria such as AIC and BIC as in Leroux (1992) and Rydén (1995). Both approaches require the evaluation of the likelihood function, which is not a easy task (Nadas and Mercer 1994, Boldin 1996). Since the sample autocovariances of time series are easily calculated, the results of the previous section can be very helpful in practice.

To illustrate, consider model selection for the mean-variance switching model:

$$y_t = \mu_{s_t} + \sigma_{s_t} \varepsilon_t, \quad (33)$$

for which $y_t \sim f(\mu_i, \sigma_i)$ when $s_t = i$. Given the results from Section 3, we can estimate a lower bound for the number of states. The idea is simple: first calculate the sample autocovariances of the observations, estimate (p', q') of the $ARMA(p', q')$ representation from these, and then use the result in Section 3 to bound k . As a corollary to Theorem 3, we have the following:

Corollary 2 *The number of states k of the Markov chain $\{s_t\}$ in (33) satisfies $k \geq \max\{p' + 1, q' + 1\}$ where (p', q') is the order of the $ARMA$ representation of $\{y_t\}$.*

Proof: Use the fact that $p', q' \leq k - 1$ in Theorem 3. ■

The problem of determining (p, q) of an $ARMA(p, q)$ model is a well-developed area. Choi (1992) gives a thorough and modern treatment. Penalized likelihood methods like AIC and BIC are inappropriate here since the probability distribution of the data is not known; the data do not, for example, form a Gaussian $ARMA$ process. Instead, one can more simply seek to identify p and q from patterns in the estimated autocovariances. Pattern recognition methods only require the estimated autocovariances without making assumptions on the probability distribution of the data. Several pattern recognition methods are closely related. These include the R and S array method (Gray, Kelley, and McIntire 1978), the corner method (Beguin *et al.* 1980), and the 3-pattern method (Choi 1992). In essence, all three utilize the Yule-Walker equations and check whether certain elements of various arrays are zeros. Since the others can be represented as 3-pattern methods, we use this method to select the $ARMA$ order.

Table 1: Transition Matrix P and Poisson means λ for Cases A-C

Parameters	A		B		C	
P	1	0	0.8	0.2	0.8	0.2
	1	0	0.2	0.8	0.2	0.8
λ	4	4	4	12	4	6

We provide two numerical examples illustrating the choice of k from the autocovariance function. We also compare the accuracy of this simply computed method to previously computed results for AIC and BIC. The first example is taken from Rydén (1995) in which three different Poisson Markov regime switching models are used. Table 1 lists the model parameters. The first model, denoted as case A, is a one-state model; that is, y_t is an i.i.d. sequence of Poisson random variables with $\lambda = 4$. Cases B and C are two-state models and share the same transition matrix but different means. With $\lambda_1 = 4$ and $\lambda_2 = 12$, the two states of case B are more distinct than those of case C with $\lambda_1 = 4$ and $\lambda_2 = 6$. As in Rydén (1995), we simulated 100 replicates of sample sizes $N = 100, 500$, and 25 replicates with sample size $N = 5000$ for each model. Table 2 summarizes the estimated number of states given by AIC, BIC and the estimated lower bound given by the 3-pattern method. For these processes, the results of Poskitt and Chung (1996) imply that this lower bound indeed does estimate the number of regimes. For cases B and C, it is easy to calculate the true autocovariances and see that they both have an $ARMA(1, 1)$ representation.

The simulation results in Table 2 show that \hat{k} based on autocovariances is almost as good

Table 2: Simulation I: Comparison of AIC, BIC and pattern method (The results for AIC and BIC are taken from Rydén (1995).)

N	\hat{k}	Case A			Case B			Case C		
		AIC	BIC	Pattern	AIC	BIC	Pattern	AIC	BIC	Pattern
100	1	91	79	95	0	0	0	62	45	44
	2	9	21	3	92	68	85	37	51	44
	3	0	0	2	8	32	15	1	4	12
500	1	92	98	97	0	0	0	9	15	0
	2	8	3	2	93	99	91	90	84	71
	3	0	0	0	7	1	9	1	1	29
5000	1	24	25	25	0	0	0	0	0	0
	2	1	0	0	23	25	25	24	25	24
	3	0	0	0	2	0	0	1	0	1

as those using AIC or BIC. Except for case C, the number of times pattern identification correctly chooses k is comparable to those obtained by AIC and BIC. With $N = 100$, all three methods frequently fail to estimate the order correctly. This implementation of BIC has a tendency to overestimate the order for $N = 100$, except in case C, where all three methods underestimate k about half of the time. When $N = 500, 5000$, BIC is the best overall.

The second simulation involves normal models. Table 3 lists the model parameters. The

standard deviation is 1 for all three models. The first model, denoted as case D, is a two-state mixture model, i.e., y_t is an independent sequence observations drawn from a mixture of two normal distributions with $\mu_1 = 4$ and $\mu_2 = 1$. Cases E and F are two-state models with equal transition matrices but different means. Unlike cases B and C, cases E and F are more likely to switch than not. With $\mu_1 = 4$ and $\mu_2 = 1$, the two states of case E are more different than those of case F with $\mu_1 = 2$ and $\mu_2 = 1$.

Table 3: Transition matrix P and regime means for cases D-F, $Y_t \sim N(\mu_j, 1)$

Parameters	D		E		F	
P	0.5	0.5	0.3	0.7	0.3	0.7
	0.5	0.5	0.8	0.2	0.8	0.2
μ	4	1	4	1	2	1

For the second example, we show results for 100 replicates with sample sizes $N = 100, 500$ and 5000 in Table 4. The results are similar to those of the previous examples except for Case D, a mixture of two normals. Pattern recognition correctly indicates for about 95% of the realizations that there is no autocorrelation, and thus implies the trivial lower bound of at least one state, $k \geq 1$. In this case, the transition matrix P is singular, so the results of Poskitt and Chung (1996) do not apply; \hat{k} is only a lower bound for the number of regimes. For case F, which has two means separated by a single standard deviation, all three methods fail to estimate the order correctly in more than half of the realizations with small sample size $N = 100$; however, pattern identification does slightly better than AIC and BIC. When

sample size increases to $N = 500$, BIC outperforms the other methods.

We may draw several conclusions from these experiments. When trying to fit a Markov regime switching model, the autocovariance structure can easily be used to ascertain a lower bound on the number of regimes k . If there is no autocorrelation present, use a mixture model without Markov regime switching. If certain $ARMA(p, q)$ structures are detected, use the estimated orders to set a lower bound on the number of regimes as suggested in Corollary 2. As the $ARMA$ representation is also true for Markov regime switching autoregressive models we can, in principle, place lower bounds on the number of regimes and the order for those models as well.

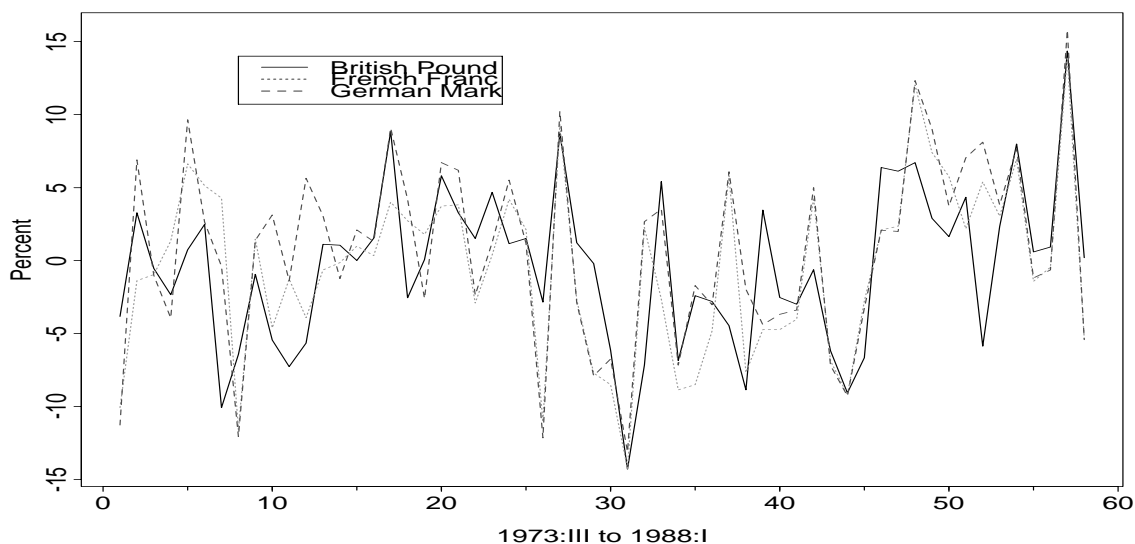
Table 4: Simulation II: Comparison of AIC, BIC and pattern method

N	\hat{k}	Case D			Case E			Case F		
		AIC	BIC	Pattern	AIC	BIC	Pattern	AIC	BIC	Pattern
100	1	2	31	93	0	2	4	64	97	55
	2	84	69	6	87	98	79	29	3	41
	3	14	0	1	13	0	17	7	0	4
500	1	0	0	96	0	0	0	7	10	24
	2	98	100	3	96	100	90	88	89	56
	3	2	0	1	4	0	10	5	1	20
5000	1	0	0	97	0	0	0	0	2	1
	2	97	100	3	100	100	91	100	98	89
	3	3	0	0	0	0	9	0	0	10

5 Example: Modeling Foreign Exchange Rates

As an illustration, we apply the 3-pattern method to several economic series and compare the results with those of other approaches. Engel and Hamilton (1990) propose a Markov regime switching model for exchange rates in which changes in the log of the exchange rate $y_t = \log e_t - \log e_{t-1}$ are modeled as a two-state Markov mean-variance switching model as defined in (33) with $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1)$. They find that this model generates better forecasts than treating y_t as a random walk, and reject the random walk model by comparing likelihoods. Here we apply pattern recognition methods to the autocovariances to estimate the number of regimes.

Figure 1: Changes of the log of the exchange rate



We use the same data series as used in Engel and Hamilton (1990); these are plotted in Figure 1. The figure shows $\log e_t/e_{t-1}$ for three currencies: the German mark, the French franc, and the British pound. All three span the 58 quarters from 1973:3 through 1988:4.

Table 5: Sample correlations, tests, and estimates of the number of regimes for the currency exchange series. The column LB (LM) gives the p-value for the test of Ljung and Box (1978) (Lo and MacKinlay 1988).

Currency	Autocorrelations				Dependence Tests		Number of Regimes		
	$\hat{\rho}(1)$	$\hat{\rho}(2)$	$\hat{\rho}(3)$	$\hat{\rho}(4)$	LB	LM	3-Pattern	AIC	BIC
Franc	0.19	0.16	0.17	0.13	0.132	0.011	≥ 2	2	1
Pound	0.28	0.07	0.22	0.10	0.045	0.004	≥ 2	2	1
Mark	0.06	0.01	0.24	0.06	0.263	0.149	≥ 2	1	1

The estimated correlations with two tests of the white noise conjecture and estimates of the number of regimes appear in Table 5. The observed correlations are small, and the Ljung-Box test for white noise (Ljung and Box 1978) only rejects the white noise hypothesis for the British pound. This test has relatively low power in problems with changing variance, so we also computed the variance ratio test of Lo and MacKinlay (1988). This test rejects the null of random walk for both the franc and pound, but not the mark. Consistent with Engel and Hamilton, application of the 3-pattern method to the estimated autocovariances indicates an ARMA(1,1) structure for each, implying a need for two (or more) regimes. When we fit Gaussian regime switching models, AIC chooses $k = 2$ for the franc and pound, and $k = 1$ for the mark. The more conservative BIC chooses a single regime, $k = 1$, for all three currencies. The two fitted regimes appear to have rather different means and variances, however. From Engel and Hamilton (1990) for the German mark, the estimated regime means (standard

error) are $\hat{\mu}_1 = 3.987$ (1.23) and $\hat{\mu}_2 = -1.183$ (1.48). The estimated variances of the two are $\hat{\sigma}_1^2 = 17.7$ (9.4) and $\hat{\sigma}_2^2 = 42.2$ (11.2) with transition probabilities $\hat{p}_{11} = 0.85$ (0.12) and $\hat{p}_{22} = 0.93$ (0.07).

6 Discussion

For a very general class of second-order stationary Markov regime switching models, the autocovariance function is that of a $VARMA(p, q)$, where the upper bound of (p, q) is an elementary function of the number of states. For a mean-variance switching process, both p and q are less than the number of regimes k . For models with switching among autoregressions, the bounds are $p \leq k(mp)^2$ and $q \leq k(mp)^2 - 1$, where k, m, p are the number of states, the dimension of the process and the autoregressive order, respectively. This result yields an easily computed method for setting a lower bound on the number of regimes from an estimated autocovariance function. Simulation results indicate the procedure is very competitive with more elaborate procedures such as AIC and BIC methods which require the assumption of a specific probability model and the associated likelihood calculations.

Since $ARMA$ models have been widely studied and successfully applied in many fields, it may be natural to ask is why use Markov regime switching models if they have an $ARMA(p, q)$ representation. The answer lies in the fact that this $ARMA(p, q)$ representation applies only to the autocovariance structure of the models. Markov switching models have much more diverse probabilistic properties than those of Gaussian $ARMA(p, q)$ processes. For example, the marginal distribution of the observations generated by Markov

regime switching model can be unimodal or bimodal as demonstrated by Rydén (1995). Furthermore, the conditional distribution of y_t given the past is different from that of a Gaussian *ARMA* process.

Acknowledgments

The first author would like to thank Professor Hall and his colleagues for the warm hospitality he received at the Australian National University where part of the research was carried out. Both authors are grateful to Professor Choi for sharing his SAS code which implements the 3-pattern method, to Professor Hamilton for providing the exchange rate data used in the illustration of Section 5, and to Professor Poskitt for the helpful discussion on Section 3. We also appreciate the comments of the editor and a very careful and articulate referee.

Appendix: Proof of Theorem 1

As in Beguin *et al.* (1980), the proof is carried out in two steps. First we establish the following

Lemma 2 *A zero mean stationary m -dimensional nondeterministic process $\{Y_t\}$ has a vector $ARMA(p, q)$ (not necessarily minimal) representation if and only if its autocovariance sequence $\gamma(h) = E(Y_t Y_{t-h}^T), h \in \mathbf{Z}$, satisfies a difference equation of order p from the rank*

$q + 1$, i.e., \exists $m \times m$ matrices $A_1, A_2, \dots, A_p, A_p \neq 0$ such that:

$$, (h) - A_1, (h - 1) \cdots - A_p, (h - p) \left\{ \begin{array}{l} = 0 \quad \text{for } h \geq q + 1 \\ \neq 0 \quad \text{for } h = q. \end{array} \right. \quad (34)$$

Proof: Necessary condition: Straight-forward calculation yields the Yule-Walker equations of (34), for example, see Brockwell and Davis (1987), p. 410.

Sufficient condition: Let $X_t = Y_t - A_1 Y_{t-1} - \cdots - A_p Y_{t-p}$. Then it is easy to see that $\{X_t\}$ is stationary and that its covariances satisfy:

$$E(X_t X_{t-h}^T) \left\{ \begin{array}{l} = 0 \quad \text{for } h \geq q + 1 \\ \neq 0 \quad \text{for } h = q. \end{array} \right. \quad (35)$$

Now we simply observe that the Wold decomposition (e.g., Hannan 1970) implies that $\{X_t\}$ has an $MA(q)$ representation,

$$X_t = \sum_{j=0}^{\infty} M_j V_{t-j}, \quad M_0 = I, \quad (36)$$

where $M_j, j \geq 0$ is a sequence of matrices and where $\{V_t\}$ forms a sequence of uncorrelated random vectors with $E(V_t V_t^T) = G$, say. From (36), we note

$$X_{t+s} = \sum_{j=0}^{s-1} M_j V_{t+s-j} + \sum_{j=s}^{\infty} M_j V_{t+s-j}, \quad (37)$$

and the two summands are uncorrelated. Since for any $s > q$ $E(X_t X_{t+s}^T) = 0$, it follows that $M_j = 0$ for all $j > q$ since V_t is uncorrelated. Thus (36) reduces to

$$X_t = \sum_0^q M_j V_{t-j}, \quad M_0 = I. \quad (38)$$

and $\{X_t\}$ has a $MA(q)$ representation as asserted. ■

Using the previous lemma, we can prove Theorem 1, which gives a necessary and sufficient condition for having the smallest (p, q) for the $VARMA(p, q)$ representation

Proof of Theorem 1: Necessary condition: We know from the necessary condition of the lemma, that if the process $\{Y_t\}$ has a $VARMA(p, q)$ representation, the sequence $\{y_t(h), h \in \mathbf{Z}\}$, satisfies a difference equation of order p from the minimal rank $q + 1$. If p were not the minimal order of this equation, there would exist a minimal order $p' < p$. From the sufficient condition of the previous lemma, $\{Y_t\}$ would have a $VARMA(p', q')$ representation with $p' < p$ and the $VARMA(p, q)$ representation would not be minimal.

Sufficient condition: We know from the sufficient condition of previous lemma that if the sequence $\{y_t(h), h \in \mathbf{Z}\}$, satisfies a difference equation of minimal order p from the minimal rank $q + 1$, the process $\{Y_t\}$ has an $VARMA(p, q)$ representation. If this representation were not minimal, there would exist a minimal $VARMA(p', q')$ representation with $p' < p$. From the necessary condition of the previous lemma, the sequence $\{y_t(h), h \in \mathbf{Z}\}$, would satisfy a difference equation of order $p' < p$ and the order p would not be minimal. ■

References

- [1] Andél, J. (1993), "A time series model with suddenly changing parameters", *J. Time Series*, **14**, 111-123.
- the 6, 121-134.

- [2] Beguin, J.M., Gouriéroux, C., and Monfort, A. (1980), “Identification of a mixed autoregressive-moving average process: the corner method”, *Time Series*, *O.D. Anderson, Ed.*, North-Holland, Amsterdam, 423-436.
- [3] Boldin, M.D. (1996), “A Check on the robustness of Hamilton’s Markov switching model approach to the economic analysis of the business cycle”, *Studies in Nonlinear Dynamics and Econometrics*, (*On-line Journal*) **1**, No. 1.
- [4] Brockwell, P.J. and Davis, R.A. (1987), *Time Series: Theory and Methods*, New York, Springer-Verlag.
- [5] Choi, B. S. (1992), *ARMA Model Identification*, New York, Springer-Verlag.
- [6] Chung, S.H., Moore, J.B., Xia, L., Premkumar, L.S. and Gages, P.W. (1990), “Characterization of single channel currents using digital signal processing techniques based on hidden Markov models,” *Phil. Trans. Roy. Soc. Lond. Ser. B*, **329**, 265-285.
- [7] Churchill, G.A. (1989), “Stochastic models for heterogeneous DNA sequences”, *Bull. Math. Biol.*, **51**, 79-94.
- [8] Donoho, D. (1988), “One-side inference about functional of a density”, *Ann. Statist.*, **16**, 1390-1420.
- [9] Engel, C. and Hamilton, J.D. (1990), “Long swings in the Dollar: Are they in the data and do markets know it?”, *Amer. Econ. Rev.*, **80**, 689-713.

- [10] Fredkin, D.R. and Rice, J.A. (1992), “Bayesian restoration of single channel patch clamp recording”, *Biometrics*, **48**, 427-448.
- [11] Gong, F. and Mariano, R.S. (1995), “Testing under non-standard conditions in frequency domain: with applications to Markov regime switching models of exchange rates and federal fund rates”, Working paper, Dept. of Economics, University of Pennsylvania.
- [12] Gray, H. L., Kelley, G. D., McIntire, D. D. (1978), “A new approach to ARMA modeling”, *Comm. Statist. B—Simulation Comput.*, **7** 1-96.
- [13] Hamilton, J.D. (1989), “ A new approach to the economic analysis of nonstationary time series and the business cycles,” *Econometrica*, **57**, 357-384.
- [14] Hamilton, J.D. (1990), “ Analysis of time series subject to changes in regime”, *J. Econometrics*, **45**, 39-70.
- [15] Hamilton, J.D. (1996), “Specification testing in Markov switching time series models”, *J. Econometrics*, **70**, 127-157.
- [16] Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.
- [17] Hansen, B.E. (1992), “The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP”, *J. App. Econometrics*, **70**, 127-157.
- [18] Hansen, B.E. (1996), “Erratum: The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP”, *J. App. Econometrics*, **11**, 195-198.

- [19] Holst, U., Lindgren, G., Holst, J. and Thuvsholmen, M. (1994), “Recursive estimation in switching autoregressions with a Markov regime”, *J. Time Series*, **15**, 489-506.
- [20] Juang, B.H. and Rabiner, L.R. (1990), “Hidden Markov models for speech recognition,” *Technometrics*, **33**, 251-272.
- [21] Karlsen, H.A. (1990), “Doubly stochastic vector $AR(1)$ processes”, Technical Report, Dept. of Mathematics, Univ. of Bergen, Norway.
- [22] Leroux, B. G. and Puterman, M.L. (1992), “Maximum-penalized-likelihood estimation for independent and Markov dependent mixture models”, *Biometrics*, **48**, 545-558.
- [23] Lütkepohl, H. (1993), *Introduction to Multiple Time Series Analysis*, Second ed., Springer-Verlag, Berlin.
- [24] Ljung, G. M. and Box, G. E. P. (1978), “On a measure of fit in time series models.” *Biometrika*, **65**, 297-303.
- [25] Lo, A. W. and MacKinlay, A. C. (1988), “Stock market prices do not follow random walks: evidence from a simple specification test”, *Review of Financial Studies*, **1**, 41 - 66.
- [26] Marcus, M. and Minc, H. (1964), *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Inc., Boston.
- [27] Nadas, A and Mercer, R.L. (1994), “Hidden Markov chains can be hard to train”, Paper presented at 230th IMS Meeting, Cleveland.

- [28] Poskitt, D.S. and Chung, S.H. (1996), “Markov chain models, time series analysis and extreme value theory,” *Adv. Appl. Prob.*, **28**, 405-425.
- [29] Rozanov, Y. A. (1967), *Stationary Random Processes*, San Francisco, Holden-Day.
- [30] Rydén, T. (1995), “Estimating the order of hidden Markov models”, *Statistics*, **26**, 345-354.
- [31] Turner, C.M., R. Startz, and C.R. Nelson (1990), “A Markov model of heteroscedasticity”, *J. Financial Economics*, **23**, 3-22.