

COMPOSITE LIKELIHOOD UNDER HIDDEN MARKOV MODEL

Jiahua Chen¹, Yi Huang¹ and Peiming Wang²

¹*University of British Columbia and* ²*Auckland University of Technology*

Abstract: This paper proposes a composite likelihood approach as an alternative to the full likelihood approach for the analysis of time series data from hidden Markov models. The proposed method requires correctly specifying only the joint density of pairs of consecutive observations. Hence, the proposed composite likelihood is algebraically simpler than the corresponding full likelihood while it retains the information on transition probabilities. The proposed maximum composite likelihood estimator with a regularization term added to the composite likelihood is consistent, asymptotically normal, and easy to implement. This estimator overcomes a difficulty in maximum likelihood estimation: both the full and composite likelihoods are unbounded when the kernel distribution is normal. Our simulation studies show that the new estimator is highly efficient and robust. We apply the method to a time series for the USD/GBP exchange rate under a two-state hidden Markov model, as suggested by Engel and Hamilton (1990). The composite likelihood approach is more robust for inference than the full likelihood.

Key words and phrases: α -mixing, EM-algorithm, equilibrium distribution, ergodicity, finite mixture model, forward-backward algorithm, regime-switching, regularization, stationary.

1. Introduction

Let $\{Y_t\}_{t=0}^{\infty}$ be a time series. A classical platform for the analysis of time series data, directly or after a transformation, is the auto-regressive-moving-average model. It postulates

$$Y_t = c + \sum_{i=1}^p \varphi_i Y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j},$$

where the ϵ_t are white noise. Statistical methods for the reliable inference of most aspects of the model are well developed (Box, Jenkins, and Reinsel (1994)).

An implicit assumption in the model is that the time series is a linear stationary process. However, time series often exhibit nonlinearity and/or nonstationarity. For instance, an economy may be in either a fast or a slow growth phase. To accommodate these characteristics, researchers (Hamilton (1988); Engel (1994)) have found that hidden Markov models (HMM) provide a simple and

sensible alternative. The HMM postulates a series of unobserved or hidden states $\{S_t\}_{t=0}^\infty$ underlying $\{Y_t\}_{t=0}^\infty$. Given $\{S_t\}_{t=0}^\infty$, $\{Y_t\}_{t=0}^\infty$ are conditionally independent and the kernel distribution, the conditional distribution of Y_t , is completely determined by the state S_t . In a typical econometrics example, S_t assumes one of two possible states representing the fast and slow growth phases/regimes of an economy respectively, and Y_t is some performance index.

Likelihood inference has been used by Engel (1994) and others to investigate various financial time series. The HMM has also been widely used in such areas as speech recognition and genome sequencing. We refer to Cappé, Moulines, and Rydén (2005) for a more complete introduction to the HMM. A full likelihood (FL) function is defined to be proportional to the joint density of $\{Y_t\}_{t=0}^T$. A highly efficient forward-backward algorithm was developed by Baum et al. (1970) to maximize the FL. Leroux (1992) and Bickel, Ritov, and Rydén (1998) showed that the maximum likelihood estimators (MLE) are consistent and asymptotically normal under HMM when its kernel distribution satisfies certain conditions.

There are challenges with the FL approach: the complex analytical form of the FL makes the maximization algorithm numerically less stable; when the kernel distribution is normal, the maximum FL value is infinite; and the effectiveness of the FL approach can be influenced by departures from the model assumptions. These shortcomings may tip the balance in favor of the simpler and likely more robust composite likelihood (CL) approach (Lindsay (1988)).

The CL has some clear advantages. We propose using the joint density of just pairs of two consecutive observations, and the resulting statistical inferences are expected to be more robust against some degree of model mis-specification. The simpler algebraic form of the CL allows an easier and more thorough asymptotic analysis, and the numerical issues associated with the CL have a more intuitive and straightforward solution. Since observations are ordered with the bulk of the serial dependence occurring in adjacent observations, pairs of consecutive observations contain useful transition information. Hence, the CL likely suffers only minor efficiency loss compared to the FL.

There is a rich literature associated with the CL. Modern information technology enables researchers to collect large quantities of complex data and, rather than modeling their joint distribution, one can construct CLs through lower-dimensional joint distributions. The CLs greatly simplify the complexity of the analysis and in many cases suffer only a mild loss in statistical efficiency. We refer to Hjort and Varin (2008) for Markov chain models, to Engle, Shephard, and Sheppard (2008) for the estimation of time-varying covariances between the returns of assets in a high-dimensional portfolio, to Ng et al. (2011) for time series models with a latent Gaussian autoregressive process, and to Pakel, Shepard, and Sheppard (2011) for GARCH panels characterized by time-varying volatility.

Comprehensive reviews of recent work can be found in Varin (2008) and Varin, Reid, and Firth (2011).

The rest of the paper is as follows. In Section 2 we give brief introductions to the HMM model and the FL and CL functions. In Section 3 we discuss CL-based estimation and its asymptotic properties. Section 4 presents Monte Carlo studies comparing the finite-sample properties of the CL and FL approaches. Section 5 reports empirical results for the analysis of the exchange rate between the U.S. dollar and the British pound. Section 6 presents our conclusions.

2. Hidden Markov Model

2.1. Hidden Markov model and the full likelihood

Let $\{S_t\}_0^\infty$ be a homogeneous Markov chain with finite state space $\{1, \dots, N\}$ and transition probability matrix $\Gamma = [p_{ij}]_{N \times N}$ where for $i, j = 1, \dots, N$, $p_{ij} = P(S_{t+1} = j \mid S_t = i)$. Let $\{Y_t\}_{t=1}^\infty$ be conditionally independent given $\{S_t\}_{t=1}^\infty$. The kernel distribution, the conditional density of Y_t given $S_t = i$, is $f(y; \theta_i)$ and it belongs to a parametric family $\{f(y; \theta) : \theta \in \Theta\}$ with Θ a subset of Euclid space. Given a set of observations $\{Y_t\}_{t=1}^T$, we investigate inference issues on Γ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^\tau$.

Here, we assume that N is known and $\theta_i \neq \theta_j$ for all $i \neq j$. The state sequence $\{S_t\}_{t=1}^T$ is unobserved. Because S_t switches from one regime (state) to another, the hidden Markov model (HMM) is called the regime-switching model in finance. Engel and Hamilton (1990) used a two-regime HMM with a normal kernel distribution to model exchange rates between the U.S. dollar and the British pound. They found that the HMM provided a better description of the data than, for instance, a simple random walk.

We use the notation $s(1:T)$ for s_1, \dots, s_T wherever appropriate. Let

$$p_{s(1:T)} = p_{s_1} \prod_{t=2}^T p_{s_{t-1}, s_t}$$

be the probability that the HMM takes state values $s(1:T)$ over the period of interest $1:T$, where $p_i = P(S_1 = i)$. The joint density function of $Y(1:T)$, regarded as the likelihood function, is

$$L(\Gamma, \boldsymbol{\theta}) = \sum_{s(1:T)} \left\{ p_{s(1:T)} \prod_{t=1}^T f(y_t; \theta_{s_t}) \right\},$$

where the summation is over all possible state sequences $s(1:T)$ of the Markov chain.

When the transition probabilities $p_{1j} = \dots = p_{Nj} = \pi_j$ for all $j = 1, \dots, N$, the state sequence consists of independent and identically distributed random variables. Therefore $p_{s(1:T)} = \prod_{t=1}^T \pi_{s_t}$ and

$$L(\Gamma, \boldsymbol{\theta}) = \prod_{t=1}^T \left\{ \sum_{j=1}^N \pi_j f(y_t; \theta_j) \right\}.$$

Consequently, Y_1, \dots, Y_T are also independent and identically distributed random variables, here from the finite mixture distribution with density $\sum_{j=1}^N \pi_j f(y; \theta_j)$.

2.2. Composite likelihood

Under HMM, the conditional joint density function of (Y_t, Y_{t+1}) is given by

$$f(y_t, y_{t+1} \mid S_t = i, S_{t+1} = j) = f(y_t; \theta_i) f(y_{t+1}; \theta_j).$$

Let $\Psi(\theta, \theta^*)$ be a bivariate distribution on $\Theta \times \Theta$ assigning equilibrium probability π_{ij} to (θ_i, θ_j) for $i, j = 1 : N$, where

$$\pi_{ij} = P(S_t = i, S_{t+1} = j) = P(S_t = i) P(S_{t+1} = j \mid S_t = i) = p_i p_{ij}.$$

At equilibrium, the unconditional joint density function is given by

$$f(y_t, y_{t+1}; \Psi) = \int f(y_t; \theta) f(y_{t+1}; \theta^*) d\Psi. \quad (2.1)$$

Marginally, (y_t, y_{t+1}) is a sample from a finite mixture with up to N^2 components.

If we “incorrectly” regard $T - 1$ pairs of consecutive observations in $y(1:T)$ as independent bivariate random variables, the resulting likelihood function is

$$\prod_{t=1}^{T-1} f(y_t, y_{t+1}; \Psi). \quad (2.2)$$

This CL is algebraically the same as the likelihood function of a finite bivariate mixture model. Chen (1998) found that adding a regularization term $C \sum_{i,j} \log(\pi_{ij})$ for some $C > 0$ to the log-likelihood improves the finite sample properties. Thus, we take the logarithm of the CL to be

$$\ell_{cl}(\Psi) = \sum_{t=1}^{T-1} \log f(y_t, y_{t+1}; \Psi) + C \sum_{i,j} \log \pi_{ij} \quad (2.3)$$

with $f(y_t, y_{t+1}; \Psi)$ given by (2.1). From now on, we use Ψ to denote both the mixing distribution and the parameter vector containing π_{ij} 's and θ_i 's; and its meaning should be clear from the context.

It will be seen from the EM-algorithm that adding a penalty with $C = 1$ enforces a lower bound $(T + N^2 - 1)^{-1}$ on π_{ij} . Although this lower bound is small and disappears when $T \rightarrow \infty$, it effectively stabilizes the fitted parameter values of (θ_i, θ_j) . In addition, the finite sample effect of the penalty improves markedly

from $C = 0$ to $C = 1$, but it remains nearly unchanged for $C = 0.5, \dots, 2$. Thus, we take $C = 1$ as in Chen (1998). The CL in (2.3) is a pseudo-likelihood because of non-independence.

In the CL setting, we use π_{ij} as the parameters since p_{ij} and π_{ij} are mutually determined: π_{ij} is, by definition, determined by p_{ij} ; and p_{ij} can be recovered from π_{ij} through $p_{ij} = \pi_{ij}/p_i = \pi_{ij}/\sum_k \pi_{ik}$.

3. Estimation

3.1. Full and composite likelihood estimators

The global maximizer of the FL is asymptotically unique, consistent, and asymptotically normal when $T \rightarrow \infty$ under *certain* conditions. These results are given by Leroux (1992), Bickel, Ritov, and Rydén (1998) and Cappé, Moulines, and Rydén (2005). Numerically, the MLEs can be computed as the iteration limit of forward-backward algorithm (Baum et al. (1970)), which is a special case of the EM algorithm in Dempster, Laird, and Rubin (1977).

A crucial condition for consistency is to be able to smoothly extend $f(x; \theta)$ to a compact space containing Θ . This condition is violated for the normal distribution. For illustration, we compute $L(\Gamma, \theta)$ for a specific pair of Γ, θ . Let $\phi(y; \mu, \sigma)$ be the normal density function with mean μ and variance σ^2 . Let $N = 2, \mu_1 = y_1, p_1 = p_2 = p_{ij} = 0.5, \mu_2 = 0$, and $\sigma_2 = 1/\sqrt{2}$, and leave σ_1 unspecified. Here $\theta = (\mu, \sigma)^\tau$ and we have $p_{s(1:T)} = 2^{-T}$ for any state sequence $s(1:T)$. For this pair of Γ and θ , as $\sigma_1 \rightarrow 0$,

$$\begin{aligned} L(\Gamma, \theta) &\geq 2^{-T} \cdot \phi(y_1; \mu_1, \sigma_1) \prod_{t=2}^T \phi(y_t; \mu_2, \sigma_2) \\ &= \sigma_1^{-1} \cdot 2^{-1/2} \cdot (4\pi)^{-T/2} \cdot \exp\left\{-\sum_{t=2}^T y_t^2\right\} \\ &\rightarrow \infty. \end{aligned} \tag{3.1}$$

Thus, there are multiple global maxima for $L(\Gamma, \theta)$ and none of them are consistent. Since $\phi(y; \mu, \sigma)$ cannot be smoothly extended to $\sigma = 0$, the consistency proof of Leroux (1992) is not applicable. A nondegenerate local maximum seems to work well but is vulnerable. This pitfall is applicable not only to the normal distribution but also to all location-scale distribution families.

The FL function closely resembles the likelihood function under a finite mixture model, which also has the unbounded-likelihood phenomenon. For a finite mixture of normal distributions, Hathaway (1985) proposed using the constraint $\max_{i,j} \sigma_i/\sigma_j < M < \infty$ to restore the consistency and asymptotic normality of the parameter estimates. However, the size of M has a direct influence on the

resulting constrained maximum likelihood estimator, and the constraint leads to additional computational issues. Adding a penalty to the log-likelihood achieves the same goal (Chen and Tan (2009)) with only an indirect influence on the value of the estimator.

For the HMM with a normal kernel, we add another regularization term and use

$$\ell_{pcl}(\Psi) = \ell_{cl}(\Psi) - T^{-1/2} \sum_{i=1}^N \left\{ \log \left(\frac{\sigma_i^2}{\hat{\sigma}_0^2} \right) + \frac{\hat{\sigma}_0^2}{\sigma_i^2} \right\}, \quad (3.2)$$

where $\hat{\sigma}_0^2$ is the sample variance of the time series. According to Chen and Tan (2009), this suffices to restore the asymptotic properties of the maximum likelihood estimators for finite normal mixture models, and it effectively improves the convergence properties of the corresponding EM-algorithm. This claim easily extends to the HMM with a normal kernel distribution. The FL may be similarly regularized so that the resulting estimator is consistent, but that is not the focus of this paper.

3.2. Numerical computation

The forward-backward algorithm of Baum et al. (1970) provides an elegant solution to the FL-based MLE, but successive multiplication in the FL can lead to underflow. A high-quality R package, *depmixS4*, is available but it occasionally crashes, likely because of the unboundedness of the FL.

In the implementation of the EM algorithm for the FL, the unobservable state variable S_t is treated as missing data, and the iterative scheme imputes the missing values of S_t in the E-step and estimates parameter values by maximizing the complete data log FL in the M-step. With $\eta_t(i) = 1$ if $S_t = i$ and 0 otherwise, and $\xi_t(i, j) = \eta_t(i)\eta_{t+1}(j)$, the complete data log FL is given by

$$\ell_{fl}^{(c)}(\Psi) = \sum_i \eta_1(i) \log p_i + \sum_{t=1}^{T-1} \sum_{i,j} \xi_t(i, j) \log p_{ij} + \sum_{t=1}^T \sum_i \eta_t(i) \log \{f(y_t; \theta_i)\}.$$

E-step: Given the observed data and the parameter estimates $\Psi^{(k)}$ from k th iteration, the missing data $\eta_t(i)$ and $\xi_t(i, j)$ are their conditional expectations. Put $a_i(1) = p_i^{(k)} f(y_1; \theta_i)$ and $b_i(T) = 1$. For $t \in 2 : T$ or $t \in 1 : (T - 1)$, compute recursively

$$a_i(t) = \sum_j a_j(t-1) p_{ji}^{(k)} f(y_t; \theta_i),$$

$$b_i(t) = \sum_j p_{ij}^{(k)} f(y_{t+1}; \theta_j) b_j(t+1)$$

and then

$$\eta_t^{(k+1)}(i) = \frac{a_i(t)b_i(t)}{\sum_j a_j(T)},$$

$$w_t^{(k+1)}(i, j) = p_{ij}^{(k)} f(y_t; \theta_j) \frac{a_i(t-1)b_j(t)}{\sum_j a_j(T)}.$$

M-step: After the E-step, maximize

$$Q(\Psi) = \sum_i \eta_1^{(k+1)}(i) \log p_i + \sum_{t=1}^{T-1} \sum_{i,j} w_t^{(k+1)}(i, j) \log p_{ij}$$

$$+ \sum_{t=1}^T \sum_i \eta_t^{(k+1)}(i) \log \{f(y_t; \theta_i)\}$$

with respect to Ψ .

There are various alternative EM algorithms for the FL for HMMs. For example, extending Mongillo and Denève (2008) and Cappé and Moulines (2009), Cappé (2011) proposed an online EM algorithm.

The numerical solution for the CL is much simpler. The $\ell_{cl}(\Psi)$ in (2.3) is algebraically identical to a log-likelihood of the mixing distribution Ψ given a set of independent and identically distributed bivariate observations (y_t, y_{t+1}) , so the numerical problem is to maximize a likelihood function under a finite mixture model. We need pay attention to the fact that the specific CL structure requires us to impute the (S_t, S_{t+1}) values in pairs.

The CL contribution of (y_t, y_{t+1}) is given by

$$\prod_{i,j} [\pi_{ij} f(y_t; \theta_i) f(y_{t+1}; \theta_j)]^{\xi_t(i,j)}.$$

Hence, the complete data log CL (with a penalty on π_{ij}) is

$$\ell_{cl}^{(c)}(\Psi) = \sum_{t=1}^{T-1} \sum_{i,j} \xi_t(i, j) \log \{ \pi_{ij} f(y_t; \theta_i) f(y_{t+1}; \theta_j) \} + \sum_{i,j} \log \pi_{ij}$$

$$= \sum_{t=1}^{T-1} \sum_{i,j} \xi_t(i, j) \log \pi_{ij} + \sum_{i,j} \log \pi_{ij}$$

$$+ \sum_{t=1}^{T-1} \sum_{i,j} \xi_t(i, j) \log \{ f(y_t; \theta_i) f(y_{t+1}; \theta_j) \}.$$

Given $\Psi^{(k)}$, we compute the conditional expectation as if (y_t, y_{t+1}) were independent samples from the bivariate finite mixture model via Bayes' formula:

$$w_t^{(k+1)}(i, j) = E\{\xi_t(i, j) \mid y_t, y_{t+1}; \Psi^{(k)}\} = \frac{\pi_{ij}^{(k)} f(y_t; \theta_i^{(k)}) f(y_{t+1}; \theta_j^{(k)})}{f(y_t, y_{t+1}; \Psi^{(k)})}.$$

Replacing $\xi_t(i, j)$ by $w_t^{(k+1)}(i, j)$ in ℓ_{cl} , we obtain the so-called Q-function in the E-step:

$$Q(\Psi|\Psi^{(k)}) = \sum_{i,j} \left\{ \sum_{t=1}^{T-1} w_t^{(k+1)}(i, j) + 1 \right\} \log \pi_{ij} + \sum_{t=1}^{T-1} \sum_{i,j} w_t^{(k+1)}(i, j) \log \{f(y_t; \theta_i) f(y_{t+1}; \theta_j)\}.$$

The equilibrium probabilities π_{ij} must satisfy some linear constrains: $\sum_{j=1}^N \pi_{ij} = \sum_{j=1}^N \pi_{ji}$ for all $i \in 1:N$ and $\sum_{i,j} \pi_{ij} = 1$. When $N = 2$, the solution to the first term in $Q(\Psi|\Psi^{(k)})$ is

$$\pi_{11}^{(k+1)} = \frac{\sum_{t=1}^{T-1} w_t^{(k+1)}(1, 1) + 1}{T + 3}, \quad \pi_{22}^{(k+1)} = \frac{\sum_{t=1}^{T-1} w_t^{(k+1)}(2, 2) + 1}{T + 3},$$

$$\pi_{12}^{(k+1)} = \pi_{21}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \{w_t^{(k+1)}(1, 2) + w_t^{(k+1)}(2, 1)\} + 2}{2(T + 3)}.$$

When $N \geq 3$, there are no explicit solutions. Numerically, the task is to maximize a convex function under N linear constrains. In our simulations, a Newton-Raphson algorithm was used.

There is often an explicit solution to the maximization with respect to θ_i of the second term in Q ; and it is f -dependent but often simple. For instance, if f is a Poisson density, then θ_i corresponds to the conditional mean of y given $S = i$ and

$$\theta_i^{(k+1)} = \frac{\sum_{t=1}^{T-1} \{[\sum_j w_t^{(k+1)}(i, j)]y_t + [\sum_j w_t^{(k+1)}(j, i)]y_{t+1}\}}{\sum_{t=1}^{T-1} \sum_j \{w_t^{(k+1)}(i, j) + w_t^{(k+1)}(j, i)\}}.$$

When f is a normal density and the composite likelihood is given by (3.2), the M-step maximizes the regularized Q-function

$$\tilde{Q}(\Psi|\Psi^{(k)}) = Q(\Psi|\Psi^{(k)}) - T^{-1/2} \sum_{i=1}^2 \left\{ \log \left(\frac{\sigma_i^2}{\hat{\sigma}_0^2} \right) + \frac{\hat{\sigma}_0^2}{\sigma_i^2} \right\}$$

with respect to $\theta = (\mu, \sigma^2)$. The update formulas for $\pi_{ij}^{(k+1)}$ and $\mu_i^{(k+1)}$ are the same as those of $\pi_{ij}^{(k+1)}$ and $\theta_i^{(k+1)}$ in the Poisson case, but

$$\sigma_i^{2(k+1)} = \frac{\sum_t \{ \sum_j w_t^{(k+1)}(i, j) (y_t - \mu_i^{(k+1)})^2 + \sum_j w_t^{(k+1)}(j, i) (y_{t+1} - \mu_i^{(k+1)})^2 \} + 2T^{-1/2} \hat{\sigma}_0^2}{\sum_t \sum_j \{ w_t^{(k+1)}(i, j) + w_t^{(k+1)}(j, i) \} + 2T^{-1/2}}.$$

Here the penalty has placed a lower bound $T^{-3/2}\hat{\sigma}_0^2$ on $\sigma_i^{2(k+1)}$, which prevents singularity.

In practice, when a likelihood is constructed, the parameter vector is estimated by one of its global maxima. Theorem 1 in the Appendix shows that the global maximum of the CL is asymptotically unique. We use multiple initial values for the above EM-iteration to locate the global maximum.

We select a sensible local maximum for FL in simulation studies to compare the performance of the FL and CL in terms of bias and mean squared error. We also examine the two estimators in terms of the in-sample goodness-of-fit and the out-of-sample prediction.

3.3. In-sample and out-of-sample performance

The in-sample fitted values of y_t are linear combinations of the form

$$\hat{y}_t = \sum_i w_{ti} \hat{\mu}_i, \tag{3.3}$$

where w_{ti} and $\hat{\mu}_i$ are the fitted conditional probability of $S_t = i$ and the regime-specific mean given the in-sample data $y(1:T)$. They are obtained by either the CL or the FL estimators. In the following, we focus on the HMM with a normal kernel.

Applying the forward-backward recursive formula, we calculate w_{ti} as follows. Let $a_i(1) = \hat{p}_i \phi(y_1; \hat{\mu}_i; \hat{\sigma}_i)$ and $b_i(T) = 1$ for $u \in 1 : N$ and for $t \in 2 : T$,

$$\begin{aligned} a_i(t) &= \sum_j \hat{p}_{ji} f(y_t; \hat{\mu}_i, \hat{\sigma}_i) a_j(t-1), \\ b_i(t-1) &= \sum_j \hat{p}_{ij} f(y_t; \hat{\mu}_j, \hat{\sigma}_j) b_j(t), \end{aligned}$$

and for $t = 1:T$, $w_{ti} = a_i(t)b_i(t)/\sum_j a_j(T)$.

As a performance measure, the in-sample mean squared error is

$$\text{MSE}_{in} = \frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2. \tag{3.4}$$

The out-of-sample predicted value at t in the external period $T + (1:T^*)$ is

$$\tilde{y}_t = \sum_i \tilde{w}_{ti} \hat{\mu}_j, \tag{3.5}$$

where \tilde{w}_{ti} is the fitted conditional probability $Pr(S_t = i \mid y_1, \dots, y_{t-1})$ given by

$$\tilde{w}_{ti} = \sum_j \hat{p}_{ji} \hat{w}_{(t-1)j} \tag{3.6}$$

Table 1. Simulation parameter settings for two-state HMM.

Mean $(\mu_1, \mu_2) =$	(-3.7, 2.6)				
Variance $(i) =$	1	2	3		
$(\sigma_1, \sigma_2) =$	(4.1, 4.5)	(2.1, 2.3)	(2.1, 6.3)		
Transition prob $(j) =$	1	2	3	4	5
$(p_{11}, p_{22}) =$	(0.90, 0.89)	(0.75, 0.75)	(0.85, 0.65)	(0.35, 0.65)	(0.40, 0.80)

with $\hat{w}_{(t-1)i} = a_i(t-1) / \sum_j a_j(t-1)$. The out-of-sample performance measure is

$$\text{MSE}_{out} = \frac{1}{T^*} \sum_{t=T+1}^{T+T^*} (\tilde{y}_t - y_t)^2. \quad (3.7)$$

4. Monte Carlo Studies

In the first stimulation study, we selected fifteen parameter settings for the two-state HMM with a normal kernel distribution, denoted (i, j) , where i and j are the choice numbers for (σ_1, σ_2) and (p_{11}, p_{22}) respectively; see Table 1.

The first setting mimics the long swing pattern in exchange rates found by Engel and Hamilton (1990). In this pattern the two exchange-rate regimes are characterized by a positive and a negative trend respectively, with a higher probability of staying in the current state than switching to the alternative state. The three settings (i, j) with $i = 1, 2, 3$ and $j = 4$ correspond to the scenario where the time series for each of the three models is a series of independent and identically distributed random variables. The three settings (i, j) with $i = 1, 2, 3$ and $j = 5$ represent the case where regime 2 dominates. The three sets of parameter values for (σ_1, σ_2) represent the extent to which the two hidden states can be identified.

In the simulation, we set the in-sample and out-of-sample sizes to $T = 150, 300$ and $T^* = 30$ respectively. The in-sample size of $T = 150$ matches a later example. We generated 1,000 sets of time series for each of the 15 settings. For each data set, we obtained the CL and FL estimates based on the in-sample data. We then computed the bias for each parameter: the absolute difference between the true parameter value and the average of the corresponding parameter estimates over all the repetitions for a setting. We computed the standard deviation of these estimates based on 1,000 repetitions for each setting. For each time series, we also computed the in-sample and out-of-sample MSE as defined in (3.4) and (3.7), respectively.

For the Monte Carlo study we wrote our own R-code for the CL method and used the CRAN R-Package `depmixS4` (Visser and Speekenbrink (2010)) for the FL method. The package crashed several times but not often enough to distort the comparison between the FL and CL. Table 2 summarizes the results for the

Table 2. Bias(standard deviation) and average in- and out-of-sample MSEs of CL and FL.

Setting -Method	Parameter						MSE	
	μ_1	μ_2	σ_1	σ_2	p_{11}	p_{22}	In	Out
(1,1)-CL	0.226 (0.98)	0.434 (1.22)	0.286 (0.55)	0.391 (0.63)	0.119 (0.10)	0.141 (0.10)	11.89	26.13
(1,1)-FL	0.075 (0.87)	0.131 (1.09)	0.142 (0.52)	0.176 (0.62)	0.021 (0.10)	0.025 (0.09)	14.95	25.78
(1,2)-CL	0.066 (1.16)	0.449 (1.36)	0.254 (0.65)	0.344 (0.74)	0.058 (0.12)	0.085 (0.12)	11.97	27.64
(1,2)-FL	0.049 (1.53)	0.324 (1.76)	0.270 (0.83)	0.277 (0.87)	0.011 (0.17)	0.025 (0.18)	14.16	27.76
(1,3)-CL	0.610 (1.00)	0.818 (1.78)	0.451 (0.62)	0.175 (0.91)	0.140 (0.14)	0.029 (0.12)	12.63	25.91
(1,3)-FL	0.432 (1.30)	0.441 (2.38)	0.455 (0.79)	0.249 (1.04)	0.078 (0.20)	0.020 (0.20)	14.30	25.94
(1,4)-CL	0.852 (1.51)	0.874 (1.35)	0.018 (0.84)	0.438 (0.84)	0.144 (0.14)	0.141 (0.14)	12.93	28.49
(1,4)-FL	0.612 (2.39)	1.006 (2.20)	0.171 (1.13)	0.553 (1.16)	0.146 (0.30)	0.137 (0.30)	15.30	28.57
(1,5)-CL	1.609 (1.62)	1.133 (1.24)	0.148 (0.89)	0.550 (0.76)	0.136 (0.13)	0.232 (0.15)	13.52	27.32
(1,5)-FL	1.183 (2.64)	1.089 (1.98)	0.053 (1.29)	0.671 (1.13)	0.145 (0.28)	0.204 (0.30)	16.01	27.40
(2,1)-CL	0.007 (0.32)	0.029 (0.37)	0.044 (0.20)	0.066 (0.26)	0.025 (0.05)	0.031 (0.06)	3.80	9.59
(2,1)-FL	0.005 (0.29)	0.003 (0.34)	0.035 (0.20)	0.036 (0.24)	0.006 (0.04)	0.008 (0.05)	3.99	9.52
(2,2)-CL	0.010 (0.38)	0.011 (0.43)	0.034 (0.26)	0.043 (0.30)	0.017 (0.07)	0.015 (0.07)	3.24	13.08
(2,2)-FL	0.004 (0.39)	0.012 (0.43)	0.032 (0.26)	0.026 (0.30)	0.004 (0.06)	0.001 (0.07)	3.35	13.05
(2,3)-CL	0.035 (0.29)	0.106 (0.70)	0.044 (0.19)	0.042 (0.42)	0.015 (0.05)	0.017 (0.09)	3.28	11.54
(2,3)-FL	0.018 (0.28)	0.157 (0.81)	0.042 (0.20)	0.004 (0.50)	0.004 (0.05)	0.002 (0.10)	3.51	11.54
(2,4)-CL	0.203 (0.81)	0.133 (0.50)	0.038 (0.48)	0.098 (0.31)	0.027 (0.11)	0.035 (0.10)	3.16	14.27
(2,4)-FL	0.248 (1.05)	0.101 (0.61)	0.059 (0.59)	0.093 (0.39)	0.024 (0.15)	0.039 (0.14)	3.55	14.29
(2,5)-CL	0.495 (1.18)	0.171 (0.42)	0.141 (0.61)	0.125 (0.27)	0.037 (0.11)	0.054 (0.11)	3.54	12.36
(2,5)-FL	0.765 (1.54)	0.153 (0.47)	0.294 (0.77)	0.181 (0.38)	0.074 (0.17)	0.056 (0.14)	4.29	12.37
(3,1)-CL	0.047 (0.36)	1.259 (1.73)	0.179 (0.48)	0.679 (1.04)	0.071 (0.07)	0.149 (0.13)	14.23	28.05
(3,1)-FL	0.013 (0.30)	0.216 (1.05)	0.004 (0.26)	0.165 (0.61)	0.011 (0.02)	0.019 (0.07)	18.16	27.47
(3,2)-CL	0.075 (0.38)	1.309 (1.98)	0.187 (0.52)	0.697 (1.08)	0.020 (0.08)	0.117 (0.14)	14.17	31.28
(3,2)-FL	0.027 (0.35)	0.452 (1.43)	0.003 (0.36)	0.256 (0.78)	0.011 (0.09)	0.039 (0.12)	16.82	31.20
(3,3)-CL	0.010 (0.26)	1.074 (2.18)	0.056 (0.30)	0.784 (1.21)	0.012 (0.06)	0.094 (0.14)	9.92	22.80
(3,3)-FL	0.012 (0.25)	0.510 (1.80)	0.001 (0.23)	0.428 (1.04)	0.001 (0.06)	0.038 (0.14)	11.03	22.79
(3,4)-CL	0.285 (0.63)	2.102 (2.18)	0.577 (0.79)	0.996 (1.12)	0.150 (0.16)	0.159 (0.16)	14.82	36.84
(3,4)-FL	0.214 (0.68)	1.284 (2.13)	0.286 (0.78)	0.595 (1.04)	0.084 (0.19)	0.107 (0.18)	17.98	36.91
(3,5)-CL	0.618 (0.90)	2.479 (2.02)	0.991 (0.98)	0.967 (1.08)	0.138 (0.14)	0.231 (0.16)	16.04	38.84
(3,5)-FL	0.548 (1.11)	1.966 (2.43)	0.712 (1.14)	0.751 (1.16)	0.119 (0.19)	0.176 (0.22)	19.53	38.92

first Monte Carlo study in terms of the bias, the standard deviation (in brackets), and the average in-sample and out-of-sample MSEs.

The two estimators have similar biases and standard deviations. From the paired-t test at the unadjusted 5% level, CL has a significantly lower standard error for estimating μ_1 , a higher bias for μ_2 and σ_2 , and a higher bias but lower standard error for p_{11} and p_{22} . None of these are apparent to the naked eye. It is, however, clear that the CL in-sample MSEs are lower. Having lower in-sample MSEs is an indication of the model flexibility of the CL, as it specifies the joint distribution of only two consecutive observations. Do not regard the lower in-sample MSEs as efficiency gain, but a model robustness property of the CL.

When the sample size increases from $T = 150$ to $T = 300$, the biases and standard deviations of both are generally smaller. The rest of the observations are similar, so we omit these details here.

In the second simulation, we generated data from four three-state HMMs with a normal kernel distribution. We set $T = 150$ and 300 but report the results only for $T = 150$. Three-state HMMs contain many more parameters, so it is cumbersome to report all the figures. We looked toward more condensed summaries and computed the total (relative) mean square errors as

$$\sum_{s=1}^N \frac{(\hat{\mu}_s - \mu_s)^2}{\sigma_s^2}, \quad \sum_{s=1}^N \frac{(\hat{\sigma}_s - \sigma_s)^2}{\sigma_s^2}, \quad \sum_{i,j=1}^N (\hat{p}_{ij} - p_{ij})^2.$$

We first computed these three values for each data set and each model. Subsequently, we computed their averages over the simulation repetitions. Our previous simulation results indicated that CL was dramatically superior to the FL in terms of these performance measurements. Due to the intuition that the FL method is the most efficient, considered three explanations: simulation error, the beneficial effect of the penalty, and small sample size. We re-designed the simulation study accordingly. The results are in Table 3.

As we had included the true parameter values as initial values in both EM-algorithms for CL and FL, the FL results improved on correction. In applications, however, we do not have true values. This imperfect simulation experiment remains informative: CL is numerically more stable.

We removed the penalty from the CL to gauge its benefit. The results of CL with and without penalty are in columns pCL-MSE and CL-MSE. Without exception, the penalty helped. Yet the unpenalized CL still outperformed FL under Models 1 and 3, while it was comparable or slightly worse than the FL under Models 2 and 4.

We increased the number of repetitions from 1,000 to 5,000 and examined the difference between the first thousand, second thousand and so on. The variations were very small so that the outcomes in Table 3 are trustworthy, barring unforeseen errors.

With these precautions taken, we are inclined to conclude that the CL with penalty is superior to the FL. This advantage may fade away as the sample size increases when the asymptotic efficiency of the FL kicks in. We carried out the simulation for $T = 300$ under the same settings but the results were similar.

We expect the penalized FL to be competitive in efficiency but need to work out the corresponding difficult theory.

The theory that we have developed is equally applicable to the HMM with Poisson or other kernel distributions. Our third simulation study compared the

Table 3. CL and FL performance under three-state HMM with normal kernel.
 Last two columns are average MSEs for μ_i , σ_i , and p_{ij}

Model	(μ_s, σ_s)	Transition matrix	pCL-MSE	CL-MSE	FL-MSE
1	$\begin{pmatrix} -6.86 & 1.5 \\ -1.50 & 1.0 \\ 2.53 & 2.0 \end{pmatrix}$	$\begin{pmatrix} 0.453 & 0.240 & 0.307 \\ 0.300 & 0.300 & 0.400 \\ 0.147 & 0.400 & 0.453 \end{pmatrix}$	$\begin{pmatrix} 0.163 \\ 0.095 \\ 0.061 \end{pmatrix}$	$\begin{pmatrix} 0.247 \\ 0.150 \\ 0.107 \end{pmatrix}$	$\begin{pmatrix} 0.457 \\ 0.232 \\ 0.142 \end{pmatrix}$
2	$\begin{pmatrix} -6.86 & 1.5 \\ -1.50 & 1.0 \\ 2.53 & 2.0 \end{pmatrix}$	$\begin{pmatrix} 0.693 & 0.307 & 0.000 \\ 0.300 & 0.300 & 0.400 \\ 0.000 & 0.247 & 0.753 \end{pmatrix}$	$\begin{pmatrix} 0.130 \\ 0.064 \\ 0.052 \end{pmatrix}$	$\begin{pmatrix} 0.140 \\ 0.077 \\ 0.068 \end{pmatrix}$	$\begin{pmatrix} 0.120 \\ 0.067 \\ 0.057 \end{pmatrix}$
3	$\begin{pmatrix} -6.86 & 3.0 \\ -1.50 & 2.0 \\ 2.53 & 3.0 \end{pmatrix}$	$\begin{pmatrix} 0.453 & 0.240 & 0.307 \\ 0.300 & 0.300 & 0.400 \\ 0.147 & 0.400 & 0.453 \end{pmatrix}$	$\begin{pmatrix} 0.480 \\ 0.162 \\ 0.122 \end{pmatrix}$	$\begin{pmatrix} 1.293 \\ 0.388 \\ 0.500 \end{pmatrix}$	$\begin{pmatrix} 1.478 \\ 0.514 \\ 0.612 \end{pmatrix}$
4	$\begin{pmatrix} -6.86 & 3.0 \\ -1.50 & 2.0 \\ 2.53 & 3.0 \end{pmatrix}$	$\begin{pmatrix} 0.693 & 0.307 & 0.000 \\ 0.300 & 0.300 & 0.400 \\ 0.000 & 0.247 & 0.753 \end{pmatrix}$	$\begin{pmatrix} 0.426 \\ 0.117 \\ 0.142 \end{pmatrix}$	$\begin{pmatrix} 0.801 \\ 0.231 \\ 0.383 \end{pmatrix}$	$\begin{pmatrix} 0.545 \\ 0.182 \\ 0.254 \end{pmatrix}$

CL and FL under a two-state HMM with a Poisson kernel. We arrived at nearly identical conclusions to those for the two-state HMM with a normal kernel: the CL and FL have similar precision in the parameter estimates, while CL gives much lower in-sample MSEs.

A fourth simulation had data from a three-state HMM with a normal kernel distribution with a two-state HMM fitted. Parameter estimations were nonsensical but the in-sample and out-of-sample MSEs remained informative. Both the CL and FL had a much higher in-sample MSE when the fitted model was incorrect. The out-of-sample MSEs using the correct or incorrect model were similar for both the CL and FL. Neither method was preferred.

5. Example

We applied both the CL and FL to a two-state HMM to analyze changes in the exchange rate between the U.S. dollar and the British pound. The data for our study are the exchange rates expressed as U.S. dollars to one British pound for the first day of each quarter from the first quarter of 1971 to the second quarter of 2011. The data are noon buying rates in New York for cable transfers payable in foreign currencies and are from the website of the Federal Reserve Bank of St. Louis. As in Engel and Hamilton (1990), we defined the quarterly change in the exchange rate to be $y_t = 100 \times (\ln(e_t) - \ln(e_{t-1}))$ where e_t and e_{t-1} are the exchange rates for quarters t and $t - 1$. The overall sample period is divided into an in-sample period from the first quarter of 1971 to the third quarter of 2008 and an out-of-sample period from the fourth quarter of 2008 to the first quarter of 2011. Engel and Hamilton (1990) used the FL under the two-state HMM to analyze quarterly changes in this exchange rate from the fourth

Table 4. Analysis of the quarterly data of the exchange rates.

In-sample Q1, 1971 to Q3, 2008; out-of-sample Q4, 2008 to Q1, 2011								
	μ_1	μ_2	σ_1	σ_2	p_{11}	p_{22}	MSE_{in}	MSE_{out}
CL	-2.838	1.204	5.519	3.847	0.642	0.791	17.84	38.67
FL	-0.762	0.455	5.974	3.038	0.983	0.963	24.20	40.68
In-sample Q4, 1973 to Q1, 1988; out-of-sample Q2, 1988 to Q3, 1990								
	μ_1	μ_2	σ_1	σ_2	p_{11}	p_{22}	MSE_{in}	MSE_{out}
CL	-4.622	2.779	4.213	4.270	0.590	0.645	10.89	24.85
FL	-3.730	2.699	4.664	4.514	0.910	0.932	18.60	27.32

quarter of 1973 to the first quarter of 1988. Their data source was different from ours.

We fitted the two-state HMM to the in-sample data of y_t using both the CL and FL methods, and calculated the in-sample and out-of-sample performance measures with $T = 151$ and $T^* = 10$. The top panel in Table 4 gives the parameter estimates and MSEs.

There are significant discrepancies in the parameter estimates between the CL and FL methods based on the in-sample data from Q1, 1971 to Q3, 2008. According to the CL, the two exchange-rate regimes have the expected quarterly changes of -2.838% and 1.204% respectively. The corresponding figures for the FL are -0.762% and 0.455% . This suggests that the FL estimator tends to underestimate both negative and positive trends in the exchange rate. Furthermore, for the CL the probability of remaining in the current regime is 0.642 for regime 1 and 0.791 for regime 2, whereas the corresponding figures for the FL are 0.983 and 0.963. Hence, the results of the CL indicate a stronger asymmetric effect in the persistence of the two exchange-rate regimes and a higher odds of regime switching.

We also examined the differences in estimated and forecast values between the CL and FL methods. Figure 1 shows the observed quarterly changes (solid line) together with the estimated and forecast quarterly changes based on the CL (dotted line) and the FL (dashed line). The values based on the CL better match the observed changes. Furthermore, we calculated the in- and out-of-sample performance measures defined by (3.4) and (3.7). The value of MSE_{in} is 17.84 for the CL and 24.20 for the FL, while the value of MSE_{out} is 38.67 for the CL and 40.68 for the FL. The CL has better in-sample and out-of-sample performance in this example.

As in Engel and Hamilton (1990), we also checked whether or not there is evidence rejecting the null hypothesis that the exchange rate follows a random walk in favor of the two-state HMM. We fitted the data of the quarterly changes to a normal distribution, and calculated both estimated and forecast values for in- and out-of-sample periods. The resulting MSE is 24.53 for the in-sample period

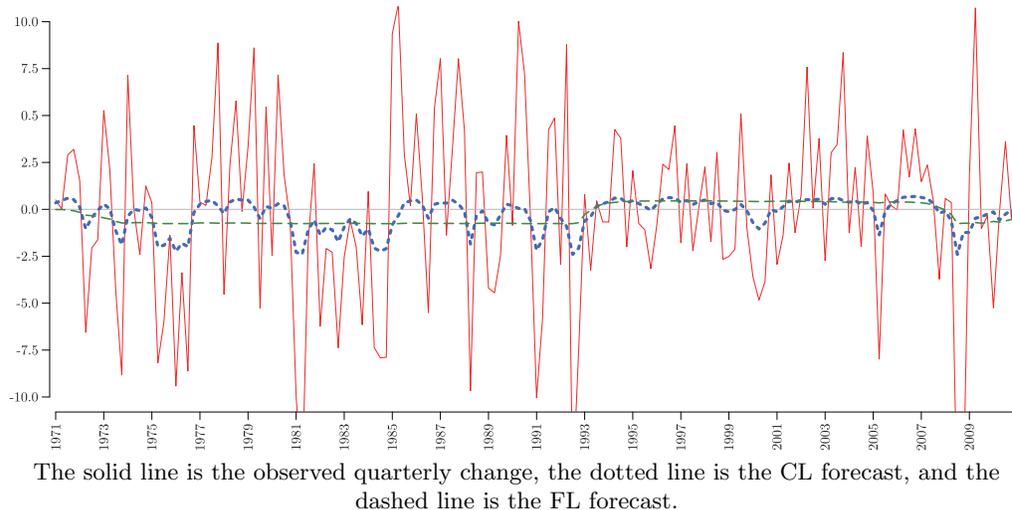


Figure 1. Observed, estimated, and forecast quarterly changes in exchange rate.

and 54.93 for the out-of-sample period. These outcomes support the observations of Engel and Hamilton (1990) that the two-state HMM is more suitable for the exchange-rate data than a random walk specification. In addition, we regressed y_{t+1} against y_t as a simple linear model and calculated the estimated and forecast values. The MSE is 24.66 for the in-sample period and 56.88 for the out-of-sample period. The results are again in favor of the two-state HMM.

We looked into the differences in the inference about the regime switching of the exchange rate between the CL and FL methods. In Figure 2, the top panel shows the observed exchange rate, and the middle and bottom panels show the estimated conditional probability $P(S_t = 1|y(1:T))$ that the exchange rate is in regime 1 for quarter t during the in-sample period based on the FL and CL methods. Econometricians consider the exchange rate to be in regime 1 if $P(S_t = 1|y(1:T)) > 0.5$ and in regime 2 otherwise. Accordingly, the shaded areas in the middle and bottom panels of Figure 2 indicate the periods of regime 1 based on the FL and CL methods respectively. Figure 2 shows that the classification of the two underlying regimes based on the CL method more closely captures the volatile periods for the exchange rate before the British government was forced to withdraw the pound sterling from the European Exchange Rate Mechanism on 16 September 1992 after it was unable to keep the currency above its agreed lower limit. Hence, the CL method characterizes the swing pattern in the exchange rate better than the FL method.

For the purposes of comparison, we repeated the above analysis for the sub-sample data with $T = 62$ and $T^* = 10$ where the in-sample period is the same as

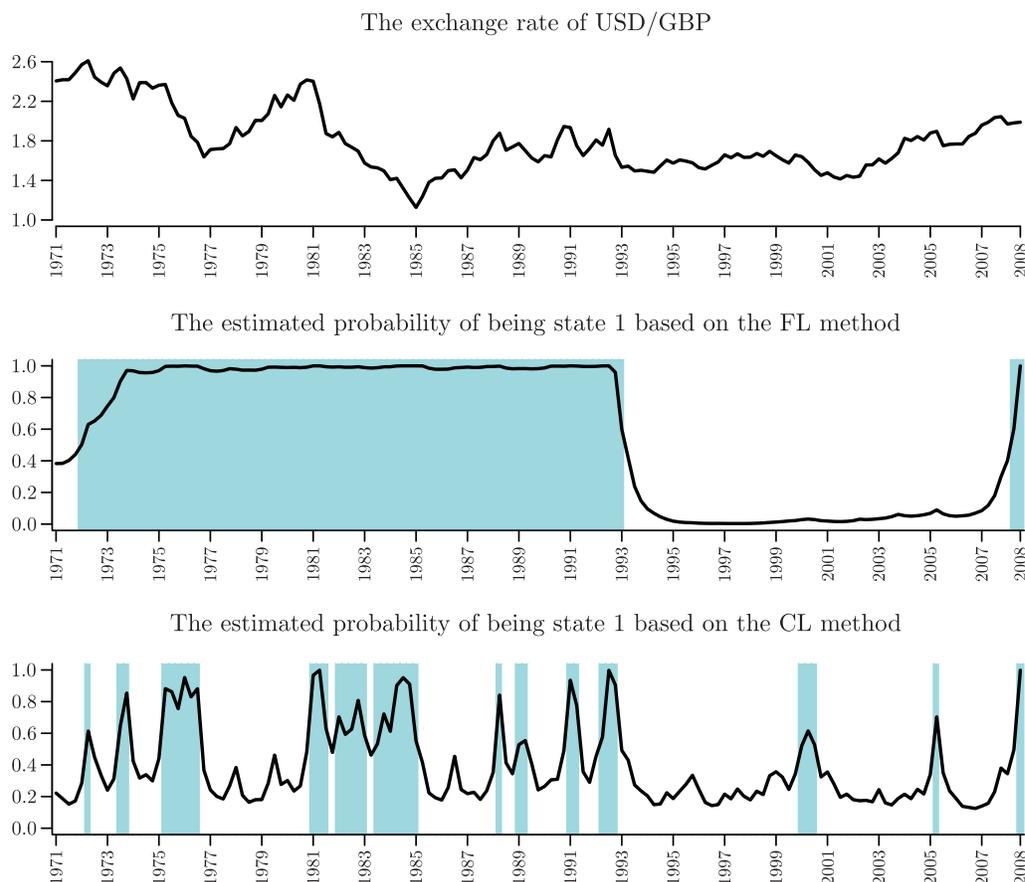


Figure 2. Observed exchange rate and estimated probabilities of being in state 1.

the sample period in Engel and Hamilton (1990). The parameter estimates and MSEs are included in the bottom panel of Table 4. The FL estimates are consistent with those of Engel and Hamilton (1990). While the FL and CL estimates of the conditional mean and standard deviation are similar, the CL estimates of the transition probabilities are much smaller, leading to more frequent regime switches. The CL method does better than the FL method in terms of both in- and out-of-sample performance: the values of MSE_{in} and MSE_{out} for the CL method are smaller. In addition, in terms of the classification of the two underlying regimes, the CL results for the two in-sample data sets are similar whereas the FL results are rather different. This suggests that the CL method is more robust for inference about regime switching.

We examined the differences in the results between the CL and FL methods when we used a three-state HMM to analyze both the whole sample and the subsample data. As in the case of the two-state HMM, the differences between

the CL and FL parameter estimates are much larger for the in-sample period from Q1, 1971 to Q3, 2008, and the CL method has better in- and out-of-sample performance since the values of MSE_{in} and MSE_{out} for the CL method are smaller for both the sample periods. We have omitted the details; and the results are available upon request.

6. Conclusion

We have proposed a CL approach for data analysis under the HMM. The CL function is constructed based on pairs of consecutive observations. We have discussed the consistency and asymptotic properties of the CL estimator. A simulation study shows that the CL has efficiency comparable to that of the FL. We have used a two-state HMM to analyze exchange rates between US dollars and British pounds. The CL classification of the periods of dollar appreciation and depreciation is more consistent and closer to the actual dollar movements. The CL estimates based on two periods of data are closer, indicating that the method has a degree of model robustness.

We have assumed that the number of regimes is known for the Markov regime-switching model. The simpler mathematical structure of the CL may make it a better choice than the FL for statistical inference on the number of underlying regimes under the Markov regime-switching model. We hope to investigate this in the future.

Acknowledgements

The research was supported by the Natural Sciences and Engineering Research Council of Canada and the UBC Killam faculty research fellowship. The authors wish to thank the Editor, and associate editor, and the referees for their rigorous review of this paper.

References

- Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximum technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.* **41**, 164-171.
- Bickel, P. J., Ritov, Y. and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* **26**, 1614-1635.
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. 3rd edition. Prentice-Hall.
- Cappé, O. (2011). Online EM algorithm for hidden Markov models. *J. Comput. Graph. Statist.* **20**, 728-749.
- Cappé, O. and Moulines, E. (2009). On-line expectation-maximization algorithm for latent data models. *J. Roy. Statist. Soc. Ser. B* **71**, 593-613.

- Cappé, O., Moulines, E. and Rydén, T. (2005). *Inference in Hidden Markov Models*. Springer.
- Chen, J. (1998). Penalized likelihood-ratio test for finite mixture models with multinomial observations. *Canad. J. Statist.* **26**, 583-599.
- Chen, J. and Tan, X. (2009). Inference for multivariate normal mixtures. *J. Multivariate Anal.* **100**, 1367-1383.
- Chen, J., Tan, X. and Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statist. Sinica* **18**, 443-465.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.
- Engel, C. (1994). Can the Markov switching model forecast exchange rates? *J. Internat. Econom.* **36**, 151-165.
- Engel, C. and Hamilton, J. D. (1990). Long swings in the dollar: Are they in the data and do the markets know it? *Amer. Econom. Rev.* **80**, 689-713.
- Engle, R. F., Shephard, N. and Sheppard, K. (2008). Fitting vast dimensional time-varying covariance models. Working Paper FIN-08-009, NYU Stern School of Business, Department of Finance.
- Hamilton, J. (1988). Rational-expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates. *J. Econom. Dynamics and Control* **12**, 385-423.
- Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Ann. Statist.* **13**, 795-800.
- Hjort, N. L. and Varin, C. (2008). ML, PL, and QL in Markov chain models. *Scand. J. Statist.* **35**, 64-82.
- Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* **40**, 127-143.
- Lindsay, B. G. (1988). Composite likelihood methods. *Contemp. Math.* **80**, 221-239.
- Mongillo, G. and Denève, S. (2008). Online learning with hidden Markov models. *Neural Comput.* **20**, 1706-1716.
- Ng, C. T., Joe, H., Karlis, D. and Liu, J. (2011). Composite likelihood for time series model with a latent autoregressive process. *Statist. Sinica* **21**, 279-305.
- Pakel, C., Shepard, N. and Sheppard, K. (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statist. Sinica* **21**, 307-329.
- Varin, C. (2008). On composite marginal likelihoods. *Adv. in Statist. Anal.* **92**, 1-28.
- Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21**, 5-42.
- Visser, I. and Speekenbrink, M. (2010). depmixS4: An R package for hidden Markov models. *J. Statist. Software* **36**, 1-21.

Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

E-mail: jhchen@stat.ubc.ca

Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

E-mail: yihuang@stat.ubc.ca

Department of Finance, Auckland University of Technology, Auckland 1142, New Zealand.

E-mail: peiming.wang@aut.ac.nz

(Received April 2013; accepted February 2014)