中央研究院經濟所學術研討論文
IEAS Working Paper

Estimating Markov-Switching ARMA Models with Extended Algorithms of Hamilton

Chao-Chun Chen and Wen-Jen Tsay

IEAS Working Paper No. 07-A009

October, 2007

Institute of Economics
Academia Sinica
Taipei 115, TAIWAN
http://www.sinica.edu.tw/econ/

中央研究院 經濟研究所

INSTITUTE OF ECONOMICS, ACADEMIA SINICA

TAIWAN

Estimating Markov-Switching ARMA Models with Extended Algorithms of Hamilton

Chao-Chun Chen and Wen-Jen Tsay*

October 3, 2007

Abstract

This paper proposes two innovative algorithms to estimate a general class of N-state Markov-switching autoregressive moving-average (MS-ARMA) models with a sample of size T. To resolve the problem of N^T possible routes induced by the presence of MA parameters, the first algorithm is built on Hamilton's (1989) method and Gray's (1996) idea of replacing the lagged error terms with their corresponding conditional expectations. We thus name it as the Hamilton-Gray (HG) algorithm. The second method refines the HG algorithm by recursively updating the conditional expectations of these errors and is named as the extended Hamilton-Gray (EHG) algorithm. The computational cost of both algorithms is very mild, because the implementation of these algorithms is very much similar to that of Hamilton (1989). The simulations show that the finite sample performance of the EHG algorithm is very satisfactory and is much better than that of the HG counterpart. We also apply the EHG algorithm to the issues of dating U.S. business cycles with the same real GNP data employed in Hamilton (1989). The turning points identified with the EHG algorithm resemble closely to those of the NBER's Business Cycle Dating Committee and confirm the robustness of the findings in Hamilton (1989) about the effectiveness of Markov-switching models in dating U.S. business cycles.

Key words: Markov-switching; ARMA process

JEL classification: C22

---

*We would like to thank the financial support from the Social Science Research Center of Taiwan. Chao-Chun Chen. The Department of Finance, Tunghai University, Taichung, Taiwan. All correspondence to Wen-Jen Tsay. The Institute of Economics, Academia Sinica, Taipei, Taiwan, R.O.C. Tel.: (886-2) 2782-2791 ext. 296. Fax: (886-2) 27853946. E-Mail: wtsay@ieas.econ.sinica.edu.tw

1 Introduction

This paper considers the possibility of estimating a general class of N-state Markov-switching autoregressive moving-average (MS-ARMA) models. The MS-ARMA(p, q) model is a natural extension to the well-known Markov-switching autoregressive (MS-AR) model proposed in the seminal paper of Hamilton (1989). It is well known that the MS-AR model performs well in modeling many macroeconomic data, especially for dating business cycles as shown in Hamilton (1989). Hamilton (1988) further explores the term structure of interest rates with the MS-AR model. See Hamilton (1994c) for other interesting macroeconomic applications of the MS-AR model.

The Markov-switching model has also been widely used in financial data. Particularly, Engel and Hamilton (1990), Engel (1994), and Bollen, et al. (2000) find Markov-switching behavior in foreign exchange data. Pagan and Schwert (1990) adopt Markov-switching models for stock returns. Hamilton and Raj (2002) provide some current reviews concerning the Markov-switching model.

The above-mentioned research studies are all based on the algorithm of Hamilton (1989). They cannot consider the potential presence of MA parameters in the data-generating process (DGP), because the possible routes of states running from time 1 to time T expand exponentially to be N^T if Hamilton's (1989) approach is adopted. Billio et al. (1999) suggest a Bayesian method based on the data augmentation principle, and Billio and Monfort (1998) adopt the partial Kalman filter and importance sampling techniques to overcome the exponential increase of routes inherent in the MS-ARMA models. The MS-ARMA model also can be estimated with the state-space approach of Kim (1994), who employs the collapsing method of Harrison and Stevens (1976) to approximate the associated likelihood function. The common feature shared with Billio et al. (1999), Billio and Monfort (1998), and Kim (1994) is that the driving force underlying the MS-ARMA process is normally distributed. By constrast, this normality assumption is not required with the method proposed in this paper.

We propose two innovative algorithms for estimating the MS-ARMA models by extending the method of Hamilton (1989), as his approach is well known and mostly used in the economic literature. To resolve the aforementioned N^T exploding regime paths problem, the first algorithm adopts the idea of Gray (1996) by replacing the lagged error terms with their corresponding conditional expectations. The rational is that the problem of N^T possible routes can be resolved with the proposed algorithm. We name this method as the Hamilton-Gray (HG) algorithm.

Similar to the method in Gray (1996), the implementation of the HG algorithm does not hinge on the specific value of the regime at time t, and the conditional expectation of the lagged error terms is recursively calculated. However, this conditional expectation is not updated even though the information set is expanded. This results in the inefficient use of information contained in data and goes strongly against the idea of Hamilton (1989) who employs the Baysian device to recursively update the conditional probability that the t-th observation was generated by regime j when the t-th observation is obtained. As a consequence, we refine the HG algorithm by proposing a recursively updating procedure to compute the conditional expectations of the lagged error terms. This algorithm is named as the extended Hamilton-Gray (EHG) algorithm and is expected to perform better than the HG counterpart in estimating the MS-ARMA model. The simulations conducted in this paper show that the finite sample performance of the EHG algorithm is satisfactory and is indeed much better than that of the HG counterpart as we predict.

The remaining parts of this paper are arranged as follows. In Section 2 we present the MS-ARMA models. Section 3 illustrates the details of the proposed algorithms. Section 4 investigates the finite sample performance of the algorithms via a Monte Carlo experiment. We apply the EHG algorithm to date U.S. business cycle turning points with the real GNP data used by Hamilton (1989) in Section 5. Section 6 provides a conclusion.

2 The MS-ARMA Model

Denote $S_t \in \{1, 2, \cdots, N\}$ as the unobserved regime at time t and s_t as the realization of S_t. The state variable s_t can assume only an integer value of $1, 2, \cdots, N$, and its transition probability matrix is:

$$\mathcal{P} \equiv \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}, \tag{1}$$

where $p_{ij} = P\left(s_t = j \mid s_{t-1} = i\right)$ and $\sum_{j=1}^{N} p_{ij} = 1$ for all i.

The MS-ARMA models considered in this paper are:

$$\Phi_{s_t}(L)(w_t - \mu_{s_t}) = \Theta_{s_t}(L)\sigma_{s_t} v_t = \Theta_{s_t}(L)\varepsilon_t, \tag{2}$$

where v_t is an independently and identically distributed (i.i.d.) white noise with $E(v_t^2) = 1$, i.e., $\varepsilon_t \sim$ i.i.d.$(0, \sigma_{s_t}^2)$, and L is the usual lag operator. We impose stationarity and invertibility constraints on the AR and MA polynominals within each regime, respectively:

$$\Phi_{s_t}(L) = 1 - \phi_{1,s_t} L - \cdots - \phi_{p,s_t} L^p, \quad \Theta_{s_t}(L) = 1 + \theta_{1,s_t} L + \cdots + \theta_{q,s_t} L^q. \tag{3}$$

These conditions are summarized in the following Assumption 1.

Assumption 1. *For each $s_t = 1, \ldots, N$, (i) The roots of the polynomial $\Phi_{s_t}(L)$ and those of $\Theta_{s_t}(L)$ in (3) are all outside the unit root circle; (ii) $\Phi_{s_t}(L)$ and $\Theta_{s_t}(L)$ share no common roots; (iii) $\sigma_{s_t} > 0$; (iv) v_τ is independent of s_t for all τ and t; and (v) $\varepsilon_t \sim i.i.d.(0, \sigma_{s_t}^2)$.*

The model in (2) subsumes the MS-AR model of Hamilton (1989). Therefore, when $q = 0$ the model can be estimated with the recursive algorithm of Hamilton (1989) based on N^{p+1} possible routes connecting w_t and its p lagged values, $w_{t-1}, \ldots, w_{t-p}$. By contrast, when $q \neq 0$ the whole past sequence, $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_1\}$, is required to extract ε_t. As a consequence, we cannot apply the algorithm of Hamilton (1989) to the MS-ARMA model without modifications, because the possible routes of states running from time 1 to T are N^T when we want to filter out the whole sequence $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_T\}$ to conduct the associated maximum likelihood estimation (MLE).

The class of MS-ARMA models in (2) with $N = 2$ has been considered in Billio et al. (1999) under the Baysian framework. However, the coverage of the model in (2) is broader than that considered in Billio et al. (1999), Billio and Monfort (1998), and Kim (1994), because we do not assume that ε_t is normally distributed as they do. Before illustrating the details of the HG and EHG algorithms for the model in (2) in the next section, let us define the notation used throughout this paper. Denote $\mathcal{W}_t \equiv (w_1, w_2, \cdots, w_t)^\top$ as a column vector containing the observations from time 1 to time t. The column vector $\alpha = (\mu_1, \cdots, \mu_N, \sigma_1, \cdots, \sigma_N, \phi_{1,1}, \cdots, \phi_{p,1}, \phi_{1,2}, \cdots, \phi_{p,2}, \cdots, \phi_{1,N}, \cdots, \phi_{p,N}, \theta_{1,1}, \cdots, \theta_{q,1}, \theta_{1,2}, \cdots, \theta_{q,2}, \cdots, \theta_{1,N}, \cdots, \theta_{q,N})^\top$ and the transition probabilities p_{ij} consist of the parameters characterizing the conditional density function (c.d.f.) of w_t. The parameters α and the transition probabilities p_{ij} are stacked into one column vector ζ.

3 Methods and Main Results

The phenomenon of N^T possible regime paths for a sample of T observations is also encountered with Gray's (1996) 2-state MS-GARCH(1,1) model. The method proposed in Gray (1996) for dealing with the MS-GARCH(1,1) model is adopted in order to modify the approach of Hamilton (1989) and to estimate the MS-ARMA considered in (2). For expositional purposes, we confine the following arguments on the assumption that the observations for the MS-GARCH(1,1) model are also w_t and the error term underlying this model is ε_t as well.

Similar to all the likelihood-based methods, the error term ε_{t-1} of the MS-GARCH(1,1) model is needed to compute the corresponding conditional variances and the associated likelihood function. To circumvent the exponentially expanding regime paths problem, Gray (1996) suggests replacing ε_{t-1} with its conditional expectation, $\widehat{\varepsilon}_{t-1|\Omega_{t-2}}$ which is computed based on the following information set:

$$\Omega_{t-2} = (\mathcal{W}_{t-2}, \Pi_{t-2}, \zeta), \qquad \text{where} \qquad \Pi_{t-2} = \widehat{\varepsilon}_{t-2|\Omega_{t-3}}. \tag{4}$$

More specifically, given the idea in page 35 of Gray (1996) and $N = 2$, the conditional expectation of ε_{t-1} is computed as:

$$\begin{aligned} \widehat{\varepsilon}_{t-1|\Omega_{t-2}} &= w_{t-1} - E[w_{t-1} \mid \Omega_{t-2}] \\ &= \sum_{i=1}^{2} P(s_{t-1} = i \mid \Omega_{t-2}) \, \left(w_{t-1} - E[w_{t-1} \mid s_{t-1} = i, \Omega_{t-2}]\right), \end{aligned} \tag{5}$$

where

$$P(s_{t-1} = i \mid \Omega_{t-2}) = \sum_{m=1}^{2} P(s_{t-1} = i \mid s_{t-2} = m) \times P(s_{t-2} = m \mid \Omega_{t-2}), \qquad \forall\, i = 1, 2,$$

and $P(s_{t-2} = m \mid \Omega_{t-2})$ denotes the inference about the probability that $s_{t-2} = m$ conditional on Ω_{t-2}. Note that the conditional expectation $\widehat{\varepsilon}_{t-1|\Omega_{t-2}}$ in (5) is recursively calculated and its value is not updated even when the information set is expanded. This idea not only prevents the occurrence of N^T possible routes from estimating the MS-GARCH(1,1) model, but also leads us to modify the algorithm of Hamilton (1989) to deal with the MS-ARMA model. We thus name this modified Hamilton algorithm as the Hamilton-Gray (HG) algorithm. The details of implementing the HG algorithm will be demonstrated later.

The extended HG (EHG) algorithm also employs Gray's (1996) method of replacing ε_{t-1} with its conditional expectation to estimate the MS-ARMA(p, q) model, but with a major modification. Note that the calculation of $\widehat{\varepsilon}_{t-1|\Omega_{t-2}}$ in (5) does not depend on the value of s_t and is not updated when the information set is renewed. This incurs inefficient use of information contained in data and is very much different from the methodology of Hamilton (1989) who employs the Baysian device to recursively update the conditional probability that the t-th observation was generated by regime j when the t-th observation is obtained. The EHG algorithm is designed to recursively update the conditional expectation of the lagged error terms ε_{t-k}, $k = 1, \ldots, q$, under the MS-ARMA(p, q) scenario. This is the first reason why we expect the EHG algorithm to perform better than the HG counterpart does.

There is one more possibility to improve the HG algorithm, i.e., at time t with Ω_{t-1} at hand the value of $P(s_{t-1} = i \mid \Omega_{t-2})$ in (5) should be replaced with $P(s_{t-1} = i \mid \Omega_{t-1})$ whenever it can be estimated. The EHG algorithm is capable of embedding the information in $P(s_{t-1} = i \mid \Omega_{t-1})$ into the estimation of the model in (2). This is the second reason that the finite sample performance of the EHG algorithm shown in Section 4 is much better than that of the HG algorithm in estimating the MS-ARMA model.

3.1 EHG algorithm

We first explain the implementation of the EHG algorithm in that the HG algorithm can be viewed as a special case of the EHG algorithm. Their difference hinges on the method of calculating the conditional expectations of the lagged error terms. For expositional purposes, we assume ε_t is normally distributed in this section, i.e., $\varepsilon_t \sim i.i.d.\ N(0, \sigma^2_{s_t})$ although our methods do not need to impose this restrictive assumption.

Denote $l = \text{Max}(p, q)$ and define a state variable s^*_t to characterize the regime path from time t to $t-l$ as follows:

$$\begin{cases} s^*_t = 1, & \text{if} \quad s_t = 1,\ s_{t-1} = 1, \cdots,\ \text{and } s_{t-l} = 1; \\ s^*_t = 2, & \text{if} \quad s_t = 2,\ s_{t-1} = 1, \cdots,\ \text{and } s_{t-l} = 1; \\ \quad \vdots & \qquad\qquad \vdots \\ s^*_t = N^{l+1}, & \text{if} \quad s_t = N,\ s_{t-1} = N, \cdots,\ \text{and } s_{t-l} = N. \end{cases} \tag{6}$$

The $(N^{l+1} \times N^{l+1})$ transition probability matrix of s^*_t, $\mathcal{P}^*$, is composed of the transition probabilities p_{ij} in (1) as follows:

$$\mathcal{P}^* \equiv \begin{bmatrix} p^*_{11} & p^*_{21} & \cdots & p^*_{N^{l+1}1} \\ p^*_{12} & p^*_{22} & \cdots & p^*_{N^{l+1}2} \\ \vdots & \vdots & \ddots & \vdots \\ p^*_{1N^{l+1}} & p^*_{2N^{l+1}} & \cdots & p^*_{N^{l+1}N^{l+1}} \end{bmatrix}, \tag{7}$$

where $p^*_{ij} = P(s^*_t = j \mid s^*_{t-1} = i)$. In other words, we do not trace the whole past history of w_t to extract ε_t to conduct the MLE. Instead, we only trace up to l lagged observations of w_t to compute the conditional expectation of the associated lagged error terms. The choice of $l = \text{Max}(p, q)$ is to ensure that we have enough observations to compute these q conditional expectations. The accuracy of our approximation method can be improved with a larger value of l. For example, $l = \text{Max}(p, q) = 4$ for an MS-ARMA(4,2) model, but we may use $l = 5$ or other larger values to implement the estimation procedure. The method of choosing l allows us to deal with the N^{l+1} possible regime paths based on the recursive algorithm of Hamilton (1989). See Hamilton (1994b, p.3069) for the illustrations of s^* and $\mathcal{P}^*$ under the set-up, $N = p = 2$ and $q = 0$.

As shown previously, we cannot exactly extract ε_t to conduct the MLE given that we only trace up to l lagged observations of w_t. Our strategy is to follow the idea of Gray (1996) by replacing $\{\varepsilon_t, \ldots, \varepsilon_{t-q+1}\}$ with their corresponding conditional expectations. However, as compared to the information set in (4), the one used for our EHG algorithm is much more informative:

$$\Omega_t^\dagger \equiv (\mathcal{W}_t, \Pi_t^\dagger, \zeta), \quad \Pi_t^\dagger = \begin{bmatrix} \widehat{\varepsilon}_{t|s_t^*=1,\Omega_{t-1}^\dagger} & \widehat{\varepsilon}_{t|s_t^*=2,\Omega_{t-1}^\dagger} & \cdots & \widehat{\varepsilon}_{t|s_t^*=N^{l+1},\Omega_{t-1}^\dagger} \\ \widehat{\varepsilon}_{t-1|s_t^*=1,\Omega_{t-1}^\dagger} & \widehat{\varepsilon}_{t-1|s_t^*=2,\Omega_{t-1}^\dagger} & \cdots & \widehat{\varepsilon}_{t-1|s_t^*=N^{l+1},\Omega_{t-1}^\dagger} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\varepsilon}_{t-q+1|s_t^*=1,\Omega_{t-1}^\dagger} & \widehat{\varepsilon}_{t-q+1|s_t^*=2,\Omega_{t-1}^\dagger} & \cdots & \widehat{\varepsilon}_{t-q+1|s_t^*=N^{l+1},\Omega_{t-1}^\dagger} \end{bmatrix}, \quad (8)$$

where the matrix $\Pi_t^\dagger$ contains the conditional expectation of the sequence $\{\varepsilon_t, \ldots, \varepsilon_{t-q+1}\}$ based on the path consistent with regime $s_t^* = j$ $(j = 1, 2, \cdots, N^{l+1})$ and the information set $\Omega_{t-1}^\dagger$. Each column in $\Pi_t^\dagger$ represents these conditional expectations under a specific value of s_t^*. The information set in (8) implies that the conditional expectation of the sequence $\{\varepsilon_t, \ldots, \varepsilon_{t-q+1}\}$ is updated as a new observation arrives and is in sharp contrast with the one displayed in (4) which is not recurively renewed.

For the calculation of $\Pi_t^\dagger$ in (8), we first note that the value of s_t^* in (6) represents a sequence of states $\{s_t, s_{t-1}, \cdots, s_{t-l}\}$. We then define $s_{t-k}(s_t^* = j)$ as the value of s_{t-k} when the regime s_t^* is j. Adopting the idea of Gray (1996) displayed in (5), the value of $\widehat{\varepsilon}_{t|s_t^*=j,\Omega_{t-1}^\dagger}$ in (8) can be calculated recursively as:

$$\begin{aligned} \widehat{\varepsilon}_{t|s_t^*=j,\Omega_{t-1}^\dagger} &= w_t - E\left(w_t \mid s_t^* = j, \Omega_{t-1}^\dagger\right), \\ &= w_t - \mu_{s_t(s_t^*=j)} - \sum_{k=1}^{p} \phi_{k,s_t(s_t^*=j)} \left(w_{t-k} - \mu_{s_{t-k}(s_t^*=j)}\right) \\ &\quad - \sum_{k=1}^{q} \theta_{k,s_t(s_t^*=j)} \widehat{\varepsilon}_{t-k|s_t^*=j,\Omega_{t-1}^\dagger}, \qquad \forall\ j = 1, 2, \cdots, N^{l+1}, \end{aligned} \quad (9)$$

where

$$\begin{aligned} \widehat{\varepsilon}_{t-k|s_t^*=j,\Omega_{t-1}^\dagger} &= \frac{\sum_{i=1}^{N^{l+1}} P(s_t^* = j, s_{t-1}^* = i \mid \Omega_{t-1}^\dagger) \times \widehat{\varepsilon}_{t-k|s_{t-1}^*=i,\Omega_{t-2}^\dagger}}{P(s_t^* = j \mid \Omega_{t-1}^\dagger)}, \\ &= \frac{\sum_{i=1}^{N^{l+1}} P(s_t^* = j \mid s_{t-1}^* = i) \times P(s_{t-1}^* = i \mid \Omega_{t-1}^\dagger) \times \widehat{\varepsilon}_{t-k|s_{t-1}^*=i,\Omega_{t-2}^\dagger}}{P(s_t^* = j \mid \Omega_{t-1}^\dagger)}, \\ &\qquad \forall\ k = 1, 2, \cdots, q;\ j = 1, 2, \cdots, N^{l+1}, \end{aligned} \quad (10)$$

such that

$$P(s_t^* = j \mid \Omega_{t-1}^\dagger) = \sum_{i=1}^{N^{l+1}} P(s_t^* = j, s_{t-1}^* = i \mid \Omega_{t-1}^\dagger). \quad (11)$$

The term $P(s^*_{t-1} = i \mid \Omega^\dagger_{t-1})$ in (10) denotes the inference about the probability that $s^*_{t-1} = i$ based on the information set $\Omega^\dagger_{t-1}$. All the elements in $\Pi^\dagger_t$ can be recursively calculated by (9) and (10) provided that we have $P(s^*_{t-1} = i \mid \Omega^\dagger_{t-1})$ and $\Pi^\dagger_{t-1}$. The value of $P(s^*_{t-1} = i \mid \Omega^\dagger_{t-1})$ across i is collected into one vector, $\widehat{\xi}_{t-1|t-1}$:

$$\widehat{\xi}_{t-1|t-1} = \begin{bmatrix} P(s^*_{t-1} = 1 \mid \Omega^\dagger_{t-1}) \\ P(s^*_{t-1} = 2 \mid \Omega^\dagger_{t-1}) \\ \vdots \\ P(s^*_{t-1} = N^{l+1} \mid \Omega^\dagger_{t-1}) \end{bmatrix}. \tag{12}$$

Moreover, $\widehat{\xi}_{t|t}$ can be found by iterating on (22.5.5) and (22.4.6) of Hamilton (1994a, p. 692) as follows:

$$\widehat{\xi}_{t|t} = \frac{\widehat{\xi}_{t|t-1} \odot \eta_t}{\mathbf{1}^\top \left(\widehat{\xi}_{t|t-1} \odot \eta_t\right)}, \tag{13}$$

$$\widehat{\xi}_{t+1|t} = \mathcal{P}^* \times \widehat{\xi}_{t|t}, \tag{14}$$

where $\mathbf{1}$ represents an $(N^{l+1} \times 1)$ vector of ones, the symbol $\odot$ denotes an element-by-element multiplication, and η_t is the conditional density of w_t given s^*_t and $\Omega^\dagger_{t-1}$:

$$\eta_t = \begin{bmatrix} f\left(w_t \mid s^*_t = 1, \Omega^\dagger_{t-1}\right) \\ f\left(w_t \mid s^*_t = 2, \Omega^\dagger_{t-1}\right) \\ \vdots \\ f\left(w_t \mid s^*_t = N^{l+1}, \Omega^\dagger_{t-1}\right) \end{bmatrix}, \tag{15}$$

such that

$$f\left(w_t \mid s^*_t = j, \Omega^\dagger_{t-1}\right) = \frac{1}{\sqrt{2\pi}\sigma_{s_t(s^*_t=j)}} \exp\left\{ \frac{-\left(\widehat{\varepsilon}_{t|s^*_t=j,\Omega^\dagger_{t-1}}\right)^2}{2\sigma^2_{s_t(s^*_t=j)}} \right\},$$

$$\forall\ j = 1, 2, \cdots, N^{l+1}. \tag{16}$$

The starting value $\widehat{\xi}_{1|0}$ can be set to be the vector of unconditional probabilities described in (22.2.26) of Hamilton (1994a, p.684).

It follows that the parameters ζ can be estimated by maximizing the following log-likelihood function with respect to these unknown parameters:

$$L(\zeta) = \sum_{t=1}^{T} \log\left(f\left(w_t \mid \Omega^\dagger_{t-1}\right)\right),$$

where

$$f\left(w_t \mid \Omega_{t-1}^{\dagger}\right) = \mathbf{1}^{\top}\left(\widehat{\xi}_{t|t-1} \odot \eta_t\right).$$

See (22.4.7) and (22.4.8) of Hamilton (1994a, p. 692) for details.

With $\widehat{\xi}_{t-1|t-1}$, we have a simple formula to compute $\widehat{\varepsilon}_{t-k|s_t^*=j,\Omega_{t-1}^{\dagger}}$ in (10) as follows:

$$\widehat{\varepsilon}_{t-k|s_t^*=j,\Omega_{t-1}^{\dagger}} = \frac{\mathbf{e}_k \Pi_{t-1}^{\dagger} \times \left(\mathcal{P}_j^{*\top} \odot \widehat{\xi}_{t-1|t-1}\right)}{\mathcal{P}_j^* \times \widehat{\xi}_{t-1|t-1}}, \qquad \forall\ j = 1, 2, \cdots, N^{l+1}, \\ k = 1, 2, \cdots, q, \tag{17}$$

where $\mathcal{P}_j^*$ denotes the j-th row of $\mathcal{P}^*$ in (7), and $\mathbf{e}_k$ indicates the k-th row of the $(q \times q)$ identity matrix $\mathbf{I}_q$. Therefore, the computational cost of the EHG algorithm is almost identical to that of Hamilton's (1989) algorithm in that the conditional expectation of the lagged error terms can be succinctly calculated with the formula in (17).

3.2 HG algorithm

We present the details of the HG algorithm in this subsection. At time t the information set in (4) used for the HG algorithm changes to be:

$$\Omega_{t-1} = \left(\mathcal{W}_{t-1}, \Pi_{t-1}, \zeta\right), \qquad \text{where} \qquad \Pi_{t-1} = \begin{bmatrix} \widehat{\varepsilon}_{t-1|\Omega_{t-2}} \\ \vdots \\ \widehat{\varepsilon}_{t-q|\Omega_{t-q-1}} \end{bmatrix}, \tag{4$'$}$$

under the MS-ARMA(p, q) scenario. Since the value of $\widehat{\varepsilon}_{t-k|\Omega_{t-1-k}}$ $(k = 2, \cdots, q)$ in $(4')$ is determined when the $(t-k)$-th observation is obtained and is not updated even when the information set is expanded, $l = p$ is what we need to trace the path of s_t for implementing the HG algorithm. As a consequence, $\mathcal{P}^*$ is an $(N^{p+1} \times N^{p+1})$ matrix when using the HG algorithm. Following (5), the first element of Π_{t-1} in $(4')$ can be calculated as:

$$\widehat{\varepsilon}_{t-1|\Omega_{t-2}} = \begin{bmatrix} \widehat{\varepsilon}_{t-1|s_{t-1}^*=1,\Omega_{t-2}} \\ \vdots \\ \widehat{\varepsilon}_{t-1|s_{t-1}^*=N^{p+1},\Omega_{t-2}} \end{bmatrix}^{\top} \times \left(\mathcal{P}^* \times \widetilde{\xi}_{t-2|t-2}\right), \tag{5$'$}$$

where $\widetilde{\xi}_{t-2|t-2}$ denotes the $(N^{p+1} \times 1)$ vector which collects conditional probabilities $P(s^*_{t-2} = m \mid \Omega_{t-2})$ for $m = 1, 2, \cdots, N^{p+1}$ and will be clearly defined in the following $(12')$. Moreover, following (9), the value of $\widehat{\varepsilon}_{t-1|s^*_{t-1}=i,\Omega_{t-2}}$ in $(5')$ is calculated as:

$$\begin{aligned}\widehat{\varepsilon}_{t-1|s^*_{t-1}=i,\Omega_{t-2}} &= w_{t-1} - E\left(w_{t-1} \mid s^*_{t-1} = i, \Omega_{t-2}\right), \\ &= w_{t-1} - \mu_{s_{t-1}(s^*_{t-1}=i)} - \sum_{k=1}^{p} \phi_{k,s_{t-1}(s^*_{t-1}=i)} \left(w_{t-1-k} - \mu_{s_{t-1-k}(s^*_{t-1}=i)}\right) \\ &\quad - \sum_{k=1}^{q} \theta_{k,s_{t-1}(s^*_{t-1}=i)} \widehat{\varepsilon}_{t-1-k|\Omega_{t-2-k}}, \qquad \forall\ i = 1, 2, \cdots, N^{p+1}. \end{aligned} \tag{9'}$$

We need to calculate the value of $\widetilde{\xi}_{t-2|t-2}$ to complete the computation of $\widehat{\varepsilon}_{t-1|\Omega_{t-2}}$ in $(5')$. Indeed, $\widehat{\xi}_{t-2|t-2}$ in (12) reduces to be $\widetilde{\xi}_{t-2|t-2}$ when the information set changes to be Ω_{t-2} and the number of regime paths s^*_{t-2} becomes N^{p+1}, i.e.:

$$\widetilde{\xi}_{t-2|t-2} = \begin{bmatrix} P(s^*_{t-2} = 1 \mid \Omega_{t-2}) \\ P(s^*_{t-2} = 2 \mid \Omega_{t-2}) \\ \vdots \\ P(s^*_{t-2} = N^{p+1} \mid \Omega_{t-2}) \end{bmatrix}. \tag{12'}$$

The value of $\widetilde{\xi}_{t|t}$ also can be found by iterating on (13) and (14) with the following $\widetilde{\eta}_t$:

$$\widetilde{\eta}_t = \begin{bmatrix} f\left(w_t \mid s^*_t = 1, \Omega_{t-1}\right) \\ f\left(w_t \mid s^*_t = 2, \Omega_{t-1}\right) \\ \vdots \\ f\left(w_t \mid s^*_t = N^{p+1}, \Omega_{t-1}\right) \end{bmatrix}, \tag{15'}$$

where

$$f\left(w_t \mid s^*_t = j, \Omega_{t-1}\right) = \frac{1}{\sqrt{2\pi}\sigma_{s_t(s^*_t=j)}} \exp\left\{\frac{-\left(\widehat{\varepsilon}_{t|s^*_t=j,\Omega_{t-1}}\right)^2}{2\sigma^2_{s_t(s^*_t=j)}}\right\} \qquad \forall\ j = 1, 2, \cdots, N^{p+1}. \tag{16'}$$

It is now clear that the HG and EHG algorithms are identical to that of Hamilton (1989) if $q = 0$. When $q \neq 0$, the EHG algorithm recursively updates the conditional expectation of the lagged error term as shown in (17). This is the first notable feature of the EHG algorithm that cannot be found in Gray's (1996) and the HG algorithms.

Another interesting feature of the EHG algorithm is that, as compared to Gray (1996) who uses $P(s_{t-1} = i \mid \Omega_{t-2})$ to calculate the value of $\widehat{\varepsilon}_{t-1|\Omega_{t-2}}$ in (5), we instead employ the value of $P(s^*_{t-1} = i \mid \Omega^\dagger_{t-1})$ for computing the value of $\widehat{\varepsilon}_{t-k|s^*_t=j,\Omega^\dagger_{t-1}}$ in (10). Since s^*_t contains the information about s_t, the EHG method is more efficient than the HG algorithm in processing the information in the data. Combining the preceding arguments, we expect the performance of the EHG algorithm to be much better than that of the HG counterpart. This is indeed what we observe in the simulation results contained in the next section.

4 Monte Carlo Experiment

We now consider the finite sample performance of the HG and EHG algorithms via a Monte Carlo experiment. We focus on the following 2-state MS-ARMA(1,1) model:

$$w_t = \mu_{s_t} + \phi_1 \left(w_{t-1} - \mu_{s_{t-1}} \right) + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \qquad \varepsilon_t \sim i.i.d.\ N(0, \sigma^2_{s_t}). \tag{18}$$

The parameters employed throughout this section are as follows:

$$\left(\sigma_1^2, \sigma_2^2 \right) = (1, 1.5), \tag{19a}$$

$$(\mu_1, \mu_2) = (1, 5), \tag{19b}$$

$$\phi_1 \in \{0.6, 0.9\}, \tag{19c}$$

$$(p_{11}, p_{22}) \in \{(0.95, 0.95), (0.5, 0.5)\}, \tag{19d}$$

$$T \in \{100, 200, 400, 800\}, \tag{19e}$$

$$\theta_1 \in \{0.5, -0.5\}. \tag{19f}$$

The parameters in (19), except the ones in (19f), have been employed in Psaradakis and Sola (1998, p.377) to evaluate the finite sample performance of Hamilton's (1989) algorithm when the DGP are the MS-AR(1) processes.

All the computations conducted in this section are performed with GAUSS. Two hundred replications are conducted for each specification. For each sample size T, 200 additional values are generated in order to obtain random starting values. The true parameters are used as the initial values for the Constrained Maximum Likelihood (CML) GAUSS program.

The maximum number of iterations for each replication is 100. We confine the search of the parameters μ_1 and μ_2 within the range of $(-20, 20)$ to ensure that the resulting estimates of these parameters are not completely unreasonable. The simulation results remain intact when this range becomes $(-50, 50)$.

Define bias as the average estimated values minus the corresponding true parameter. Tables 1 and 2 show that the performance of the HG algorithm is not well, because the bias from estimating θ_1 is sizable and cannot be alleviated even when the sample size increases to be 800. On the other hand, Tables 3 and 4 clearly demonstrate the ability of the EHG algorithm to deal with the estimation of the MS-ARMA model, because the bias is very close to zero (especially when the sample size is larger) for all specifications considered in the tables. The associated root-mean-squared error (RMSE) contained in Tables 5 and 6 is also found to decrease with the increasing values of sample size. These observations together reveal the great potential of the EHG algorithm in estimating the MS-ARMA models.

Among the 8 configurations considered in Table 5 and Table 6, we note that the performance of the EHG algorithm is relatively weak under the following two settings:

$$\{p_{11} = 0.95, p_{22} = 0.95, \phi_1 = 0.6, \theta_1 = -0.5\},$$

and

$$\{p_{11} = 0.5, p_{22} = 0.5, \phi_1 = 0.6, \theta_1 = -0.5\}.$$

These phenomena can be explained by noting that the combinations of the values of ϕ_1 and θ_1 in these two cases are close to violating the identification condition required in item (ii) of Assumption 1. However, the changing pattern of bias and RMSE from estimating ϕ_1 and θ_1 reveals that the performance of the EHG algorithm is still quite well under these two settings. The performance of the EHG algorithm for estimating the MA parameter is particularly displayed with the box-plots in Figures 1-8. The above-mentioned observations are clearly borne out in these figures.

5 Dating U.S. Business Cycles

In this section we apply the EHG algorithm to U.S. quarterly real GNP data (1951:2 to 1984:4) employed by Hamilton (1989). We wish to investigate the robustness of the findings in Hamilton (1989) concerning the effectiveness of Markov-switching models in dating U.S. business cycles when the potential presence of MA parameters is taken into account. In other words, we treat the switching ARMA(4,0) model used in Hamilton (1989) as the benchmark and re-estimate these U.S. real GNP data with the following 3 MS-ARMA($4, q$) models:

$$w_t = \mu_{s_t} + \sum_{i=1}^{4} \phi_i \left(w_{t-i} - \mu_{s_{t-i}} \right) + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \quad q = 1, 2, 3, \tag{20}$$

where

$$\varepsilon_t \sim i.i.d. \;\; N(0, \sigma^2).$$

Table 7 presents the parameter estimates. Following Hamilton (1989), the standard errors of these estimates are calculated numerically. The associated business cycles dating are contained in Table 8.

As shown in Table 7, the relatively robust variables are μ_1, μ_2, p_{11}, p_{22} and σ, which remain qualitatively intact across all specifications considered in Table 7. In addition, none of the estimates of the parameters, θ_1, θ_2, and θ_3 are statistically significant, indicating that the influence of the MA parameters on shaping the time series behaviors of the real GNP data is negligible, once we have incorporated an AR(4) polynomial into our MS-ARMA model.

For various values of q considered in Table 8, we identify 7 business cycles as found in the report of the Business Cycle Dating Committee and the results from Hamilton (1989) based on the MS-AR(4) Model. We also find that the turning points specified by the EHG algorithm in Table 8 resemble closely to those of the NBER's Business Cycle Dating Committee. The sum of the absolute dating error against the NBER dating points shown in Table 8 indicates that the MS-AR(4) and MS-ARMA(4,1) models perform equally well in dating business cycles. Indeed, the turning points associated with the MS-ARMA(4,1) model are almost identical to those found in Hamilton (1989) except for the 2nd peak and 5th peak. These observations confirm the robustness of the findings in Hamilton (1989) and indicate that the EHG algorithm is a useful method for estimating MS-ARMA models.

6 Conclusions

This paper develops two algorithms to resolve the problem of N^T exploding regime paths associated with a general class of N-state Markov-switching ARMA models based on Hamilton's approach (1989), as his method is mostly adopted in the economic literature. The EHG algorithm is particularly useful, because it processes the information contained in the data more efficiently than the HG algorithm does. The simulations confirm that the finite sample performance of the EHG algorithm is very promising and is much better than that of the HG counterpart. In addition, the computational cost of the EHG algorithm is almost identical to that of Hamilton's (1989) method, except the EHG algorithm adds a simple formula displayed in (17) in computing the conditional expectation of the lagged error terms. As a consequence, the EHG algorithm can be easily adopted by those who are familiar with Hamilton's (1989) approach and extended to estimate the multivariate MS-ARMA processes by combining the MS-VAR literature. We also apply the EHG algorithm to check the robustness of the findings in Hamilton (1989) on dating U.S. business cycles. It is found that the MS-ARMA(4,0) and MS-ARMA(4,1) models work equally well in dating U.S business cycles with the data used in Hamilton (1989). All these findings point to the potential of the EHG algorithm in estimating the MS-ARMA model which can be of great use to many empirical applications found with the existing MS-AR models.

References

1. Billio, M., Monfort, A., Robert, C.P., 1999. Bayesian estimation of switching ARMA models. Journal of Econometrics 93, 229-255.

2. Billio, M., Monfort, A., 1998. Switching state space models, likelihood function, filtering and smoothing. Journal of Statistical Planning and Inference 68, 65-103.

3. Bollen, N.P.B., Gray, S.F., Whaley, R.E., 2000. Regime switching in foreign exchange rates: Evidence from currency option prices. Journal of Econometrics 94, 239-276.

4. Engel, C., 1994. Can the Markov switching model forecast exchange rates? Journal of International Economics 36, 151-165.

5. Engel, C., Hamilton, J.D., 1990. Long swings in the dollar: Are they in the data and do markets know it? American Economic Review 80, 689-713.

6. Garcia, R., Perron P., 1996. An analysis of the real interest rate under regime shifts. Review of Economics and Statistics 78, 111-125.

7. Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.

8. Hamilton, J.D., 1988. Rational-expectations econometric analysis of change in regime: An investigation of term structure of interest rates. Journal of Economic Dynamics and Control 12, 384-423.

9. Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357-384.

10. Hamilton, J.D., 1994a. Time Series Analysis. Princeton University Press, New Jersey.

11. Hamilton, J.D., 1994b. State-space models, in: Engle R.F., McFadden, D.L., eds., Handbook of Econometrics, Vol. 4, 3039-3080. North-Holland, Amsterdam.

12. Hamilton, J.D., 1994c. Estimation, Inference and Forecasting of Time Series Subject to Changes in Regime, in: Maddala, G.S., Rao, C.R., Vinod, H.D., eds., Handbook of Statistics, Vol. 11, 3039-3080. North-Holland, Amsterdam.

13. Hamilton, J.D., Raj, B., 2002. New directions in business cycle research and financial analysis. Empirical Economics 27, 149-162.

14. Harrison, P.J., Stevens, C.F., 1976. Bayesian forecasting. Journal of the Royal Statistical Society B 38, 205-247.

15. Kim, C.J., 1994. Dynamic linear models with Markov-switching. Journal of Econometrics 60, 1-22.

16. Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267-290.

17. Psaradakis, Z., Sola, M., 1998. Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. Journal of Econometrics 86, 369-386.

Table 1. The finite sample performance of the HG algorithm: Bias

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = 0.5$							
0.95	0.95	0.6	100	0.105	0.009	0.039	-0.208	-0.005	-0.029	-0.038	-0.035
			200	0.123	-0.064	0.062	-0.231	0.021	0.022	-0.007	-0.011
			400	0.165	-0.066	0.075	-0.235	0.029	0.030	-0.005	-0.008
			800	0.153	-0.060	0.083	-0.238	0.037	0.032	-0.003	-0.006
		0.9	100	-0.036	0.209	-0.027	-0.187	-0.001	-0.024	-0.044	-0.045
			200	0.171	0.229	-0.008	-0.203	0.030	0.045	-0.019	-0.038
			400	0.140	0.146	0.005	-0.208	0.032	0.043	-0.013	-0.020
			800	0.201	0.194	0.011	-0.213	0.050	0.044	-0.002	-0.008
0.5	0.5	0.6	100	-0.054	0.119	0.092	-0.422	0.057	0.039	-0.001	-0.010
			200	-0.036	0.120	0.114	-0.415	0.071	0.054	0.002	-0.006
			400	-0.043	0.116	0.121	-0.416	0.073	0.064	-0.001	-0.006
			800	-0.041	0.115	0.124	-0.416	0.077	0.067	0.002	-0.003
		0.9	100	0.106	0.378	-0.001	-0.374	0.073	0.054	-0.005	-0.014
			200	0.140	0.389	0.012	-0.375	0.082	0.076	-0.001	-0.007
			400	0.146	0.400	0.019	-0.375	0.088	0.084	-0.004	-0.007
			800	0.138	0.386	0.023	-0.376	0.091	0.089	-0.001	-0.003

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = 0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19). Bias is defined by the mean of estimated values minus the corresponding true parameter.

Table 2. The finite sample performance of the HG algorithm: Bias

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = -0.5$							
0.95	0.95	0.6	100	-0.029	0.026	-0.498	0.473	-0.049	-0.034	-0.015	-0.013
			200	-0.010	0.015	-0.486	0.476	-0.018	-0.013	-0.006	-0.005
			400	-0.007	0.008	-0.483	0.482	-0.008	-0.003	-0.002	-0.002
			800	-0.013	0.009	-0.474	0.479	-0.004	0.001	0.000	-0.001
		0.9	100	-0.196	0.003	-0.208	0.296	-0.038	-0.004	-0.020	-0.021
			200	-0.186	0.016	-0.176	0.288	-0.004	0.037	-0.004	-0.002
			400	-0.163	0.050	-0.155	0.285	0.023	0.039	0.001	0.002
			800	-0.166	0.033	-0.140	0.283	0.026	0.045	0.003	0.003
0.5	0.5	0.6	100	0.005	-0.022	-0.502	0.490	-0.031	-0.052	-0.006	-0.014
			200	0.009	-0.007	-0.517	0.503	-0.016	-0.026	-0.003	-0.011
			400	0.001	-0.006	-0.502	0.500	-0.009	-0.011	-0.006	-0.011
			800	-0.001	-0.011	-0.492	0.500	-0.006	-0.002	-0.003	-0.005
		0.9	100	0.052	-0.113	-0.269	0.412	0.024	-0.018	-0.006	-0.024
			200	0.060	-0.096	-0.231	0.419	0.044	0.008	0.002	-0.011
			400	0.056	-0.095	-0.203	0.422	0.047	0.016	-0.001	-0.015
			800	0.056	-0.098	-0.190	0.419	0.051	0.021	0.002	-0.009

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = -0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19). Bias is defined by the mean of estimated values minus the corresponding true parameter.

Table 3. The finite sample performance of the EHG algorithm: Bias

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = 0.5$							
0.95	0.95	0.6	100	-0.074	0.035	-0.048	0.018	-0.057	-0.028	-0.020	-0.017
			200	-0.009	0.001	-0.021	0.003	-0.022	-0.006	-0.007	-0.005
			400	0.004	0.013	-0.007	-0.003	-0.007	-0.002	-0.003	-0.003
			800	-0.004	0.005	-0.002	-0.006	-0.002	0.003	0.000	-0.002
		0.9	100	-0.151	-0.048	-0.042	0.013	-0.054	-0.005	-0.023	-0.028
			200	-0.036	0.024	-0.021	0.001	-0.023	-0.004	-0.008	-0.011
			400	-0.045	-0.005	-0.010	-0.005	-0.011	0.003	-0.005	-0.002
			800	-0.045	-0.032	-0.004	-0.007	-0.003	0.006	-0.001	-0.002
0.5	0.5	0.6	100	-0.003	0.000	-0.048	0.024	-0.006	-0.034	0.002	-0.016
			200	0.003	-0.002	-0.020	0.009	0.000	-0.012	0.002	-0.009
			400	0.005	0.003	-0.009	-0.001	0.009	-0.003	0.000	-0.011
			800	0.005	-0.001	-0.005	-0.004	0.012	0.002	0.002	-0.007
		0.9	100	-0.097	-0.096	-0.036	0.007	-0.013	-0.029	0.001	-0.014
			200	-0.036	-0.041	-0.019	0.004	-0.005	-0.012	0.002	-0.008
			400	-0.014	-0.016	-0.010	0.001	0.003	-0.005	-0.001	-0.012
			800	-0.008	-0.013	-0.005	-0.004	0.007	0.000	0.001	-0.006

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = 0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19). Bias is defined by the mean of estimated values minus the corresponding true parameter.

Table 4. The finite sample performance of the EHG algorithm: Bias

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = -0.5$							
0.95	0.95	0.6	100	-0.001	0.007	-0.226	0.177	-0.049	-0.036	-0.016	-0.012
			200	0.007	-0.006	-0.187	0.167	-0.020	-0.013	-0.006	-0.004
			400	0.003	-0.002	-0.156	0.151	-0.009	-0.005	-0.002	-0.002
			800	-0.002	-0.001	-0.063	0.062	-0.005	-0.001	0.000	-0.001
		0.9	100	-0.031	0.046	-0.066	0.027	-0.050	-0.036	-0.021	-0.015
			200	0.000	0.007	-0.031	0.017	-0.020	-0.008	-0.006	-0.004
			400	-0.008	0.001	-0.014	0.010	-0.008	-0.002	-0.002	-0.002
			800	-0.009	0.002	-0.006	0.007	-0.003	0.002	0.000	-0.001
0.5	0.5	0.6	100	-0.003	-0.005	-0.284	0.242	-0.039	-0.059	-0.004	-0.013
			200	-0.012	-0.026	-0.266	0.242	-0.025	-0.021	-0.004	-0.005
			400	0.000	-0.008	-0.249	0.249	-0.013	-0.010	-0.006	-0.011
			800	-0.001	-0.011	-0.171	0.174	-0.007	-0.002	-0.003	-0.005
		0.9	100	-0.010	-0.014	-0.080	0.066	-0.013	-0.037	-0.004	-0.015
			200	-0.010	-0.019	-0.036	0.037	-0.004	-0.008	-0.001	-0.006
			400	-0.006	-0.010	-0.018	0.027	0.003	-0.002	-0.004	-0.009
			800	-0.003	-0.008	-0.009	0.018	0.006	0.002	0.000	-0.005

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = -0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19). Bias is defined by the mean of estimated values minus the corresponding true parameter.

Table 5. The finite sample performance of the EHG algorithm: RMSE

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = 0.5$							
0.95	0.95	0.6	100	0.573	0.659	0.135	0.135	0.184	0.191	0.061	0.089
			200	0.392	0.340	0.083	0.099	0.087	0.099	0.027	0.028
			400	0.257	0.259	0.051	0.060	0.060	0.070	0.019	0.020
			800	0.171	0.174	0.039	0.039	0.038	0.051	0.012	0.013
		0.9	100	1.775	1.692	0.074	0.136	0.173	0.445	0.079	0.105
			200	1.199	1.133	0.045	0.089	0.101	0.116	0.032	0.047
			400	0.891	0.890	0.028	0.056	0.064	0.072	0.028	0.021
			800	0.631	0.602	0.017	0.035	0.043	0.053	0.012	0.014
0.5	0.5	0.6	100	0.438	0.462	0.141	0.164	0.133	0.165	0.083	0.084
			200	0.283	0.310	0.084	0.098	0.078	0.106	0.057	0.063
			400	0.220	0.224	0.053	0.068	0.063	0.064	0.040	0.045
			800	0.144	0.153	0.040	0.045	0.043	0.045	0.026	0.029
		0.9	100	1.723	1.744	0.074	0.110	0.116	0.148	0.075	0.078
			200	1.118	1.140	0.041	0.075	0.070	0.098	0.055	0.059
			400	0.838	0.842	0.027	0.054	0.059	0.061	0.038	0.043
			800	0.574	0.580	0.018	0.034	0.041	0.041	0.026	0.028

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = 0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19).

Table 6. The finite sample performance of the EHG algorithm: RMSE

DGP				MLE							
p_{11}	p_{22}	ϕ_1	T	μ_1	μ_2	ϕ_1	θ_1	σ_1	σ_2	p_{11}	p_{22}
				$\theta_1 = -0.5$							
0.95	0.95	0.6	100	0.236	0.238	0.508	0.526	0.150	0.164	0.058	0.057
			200	0.147	0.146	0.476	0.483	0.078	0.092	0.027	0.026
			400	0.091	0.111	0.427	0.432	0.056	0.069	0.018	0.018
			800	0.061	0.076	0.237	0.247	0.038	0.051	0.011	0.013
		0.9	100	0.801	0.799	0.141	0.194	0.211	0.164	0.086	0.062
			200	0.444	0.424	0.076	0.104	0.093	0.096	0.027	0.026
			400	0.327	0.319	0.043	0.067	0.062	0.073	0.018	0.019
			800	0.216	0.225	0.026	0.047	0.041	0.051	0.012	0.013
0.5	0.5	0.6	100	0.258	0.324	0.518	0.566	0.182	0.233	0.103	0.106
			200	0.206	0.258	0.500	0.515	0.115	0.155	0.073	0.074
			400	0.112	0.139	0.476	0.492	0.080	0.095	0.047	0.049
			800	0.074	0.095	0.380	0.391	0.058	0.066	0.032	0.030
		0.9	100	1.187	1.198	0.172	0.237	0.158	0.221	0.085	0.100
			200	0.373	0.425	0.080	0.131	0.097	0.122	0.059	0.065
			400	0.290	0.307	0.043	0.077	0.074	0.078	0.041	0.045
			800	0.194	0.210	0.026	0.051	0.051	0.055	0.028	0.028

Notes: Simulations are based on 200 replications. The DGP is the MS-ARMA(1,1) model defined in (18) with $\theta_1 = -0.5$. Other parameters are set as $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$, as shown in (19).

Table 7. Maximum likelihood estimates of parameters and asymptotic standard errors based on data for U.S. quarterly real GNP and the EHG algorithm

	ARMA(4,0)		ARMA(4,1)		ARMA(4,2)		ARMA(4,3)	
Parameter	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
μ_1	1.164	0.074	1.173	0.073	1.174	0.073	1.176	0.076
μ_2	-0.359	0.263	-0.310	0.275	-0.312	0.226	-0.309	0.218
ϕ_1	0.013	0.116	0.265	0.348	0.372	0.391	0.167	0.424
ϕ_2	-0.058	0.137	-0.068	0.122	-0.220	0.501	0.061	0.350
ϕ_3	-0.247	0.107	-0.235	0.105	-0.206	0.149	-0.421	0.236
ϕ_4	-0.213	0.110	-0.169	0.135	-0.170	0.138	-0.161	0.142
σ	0.769	0.102	0.769	0.064	0.768	0.060	0.768	0.060
p_{11}	0.904	0.038	0.905	0.037	0.904	0.037	0.905	0.038
p_{22}	0.755	0.097	0.769	0.096	0.770	0.089	0.769	0.089
θ_1	-	-	-0.284	0.355	-0.393	0.394	-0.175	0.434
θ_2	-	-	-	-	0.170	0.560	-0.109	0.373
θ_3	-	-	-	-	-	-	0.202	0.250

Notes: The results are based on the switching ARMA$(4, q)$ model defined in (20). S.E. stands for standard error of the estimate.

Table 8. Dating U.S. business cycles with the MS-ARMA$(4, q)$ Models and the EHG algorithm

NBER		MS-ARMA(4,0)		MS-ARMA(4,1)		MS-ARMA(4,2)		MS-ARMA(4,3)	
Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough	Peak	Trough
1953:3	1954:2	1953:3	1954:2	1953:3	1954:2	1953:2	1954:2	1953:2	1954:2
1957:3	1958:2	1957:1	1958:1	1956:4	1958:1	1956:4	1958:1	1956:4	1958:1
1960:2	1961:1	1960:2	1960:4	1960:2	1960:4	1960:2	1960:4	1960:2	1960:4
1969:4	1970:4	1969:3	1970:4	1969:3	1970:4	1969:3	1970:4	1969:3	1970:4
1973:4	1975:1	1974:1	1975:1	1973:4	1975:1	1973:3	1975:1	1973:4	1975:1
1980:1	1980:3	1979:2	1980:3	1979:2	1980:3	1979:2	1980:3	1979:2	1980:3
1981:3	1982:4	1981:2	1982:4	1981:2	1982:4	1981:2	1982:4	1981:2	1982:4
		Sum of absolute dating error (quarter) against NBER							
–		10		10		12		11	

Notes: The results are based on the MS-ARMA$(4, q)$ model defined in (20) with the parameters presented in Table 7.

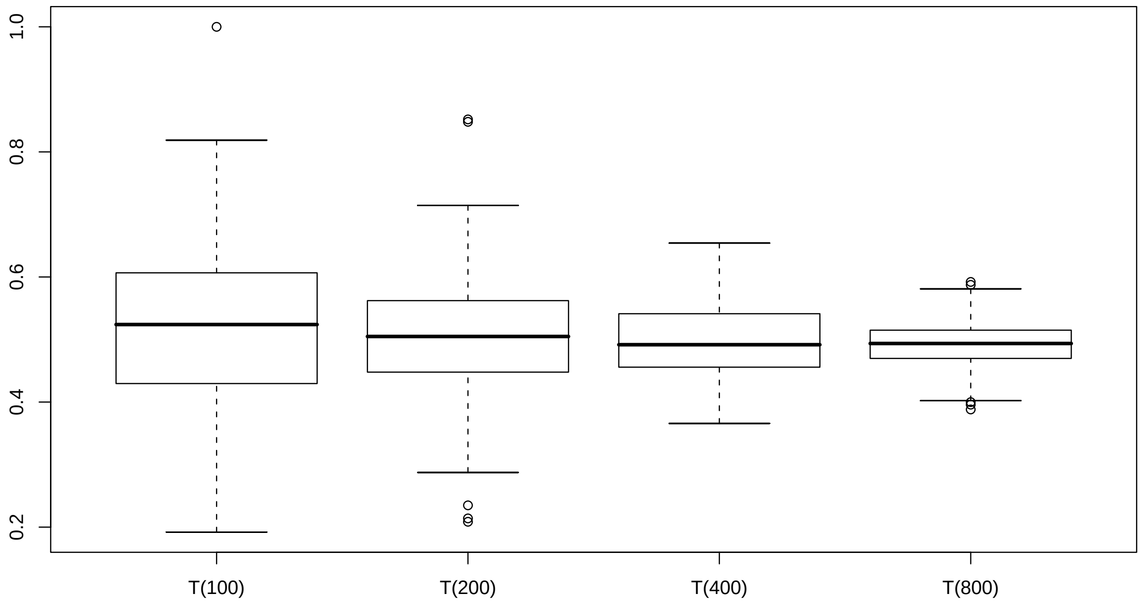

Figure 1. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.95$, $p_{22} = 0.95$, $\phi_1 = 0.6$, $\theta_1 = 0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

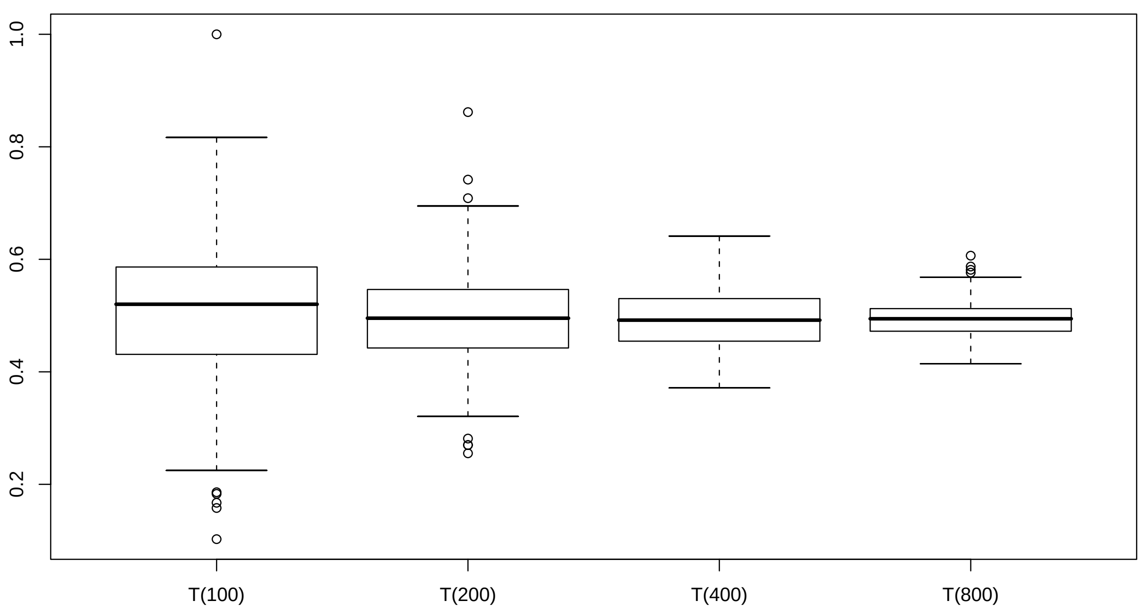

Figure 2. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.95$, $p_{22} = 0.95$, $\phi_1 = 0.9$, $\theta_1 = 0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

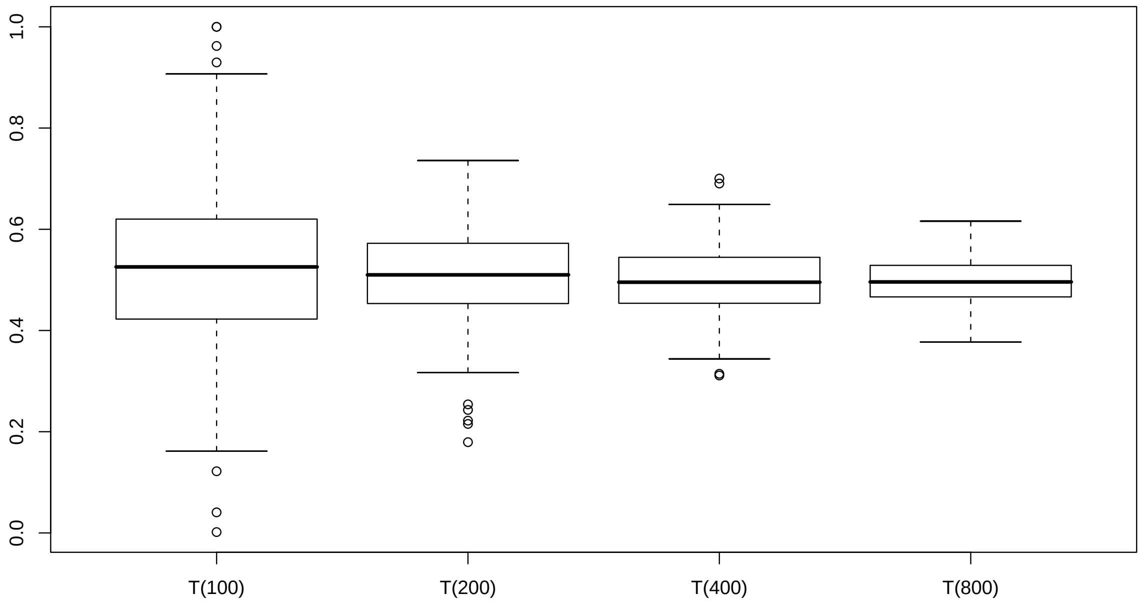

Figure 3. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.5$, $p_{22} = 0.5$, $\phi_1 = 0.6$, $\theta_1 = 0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

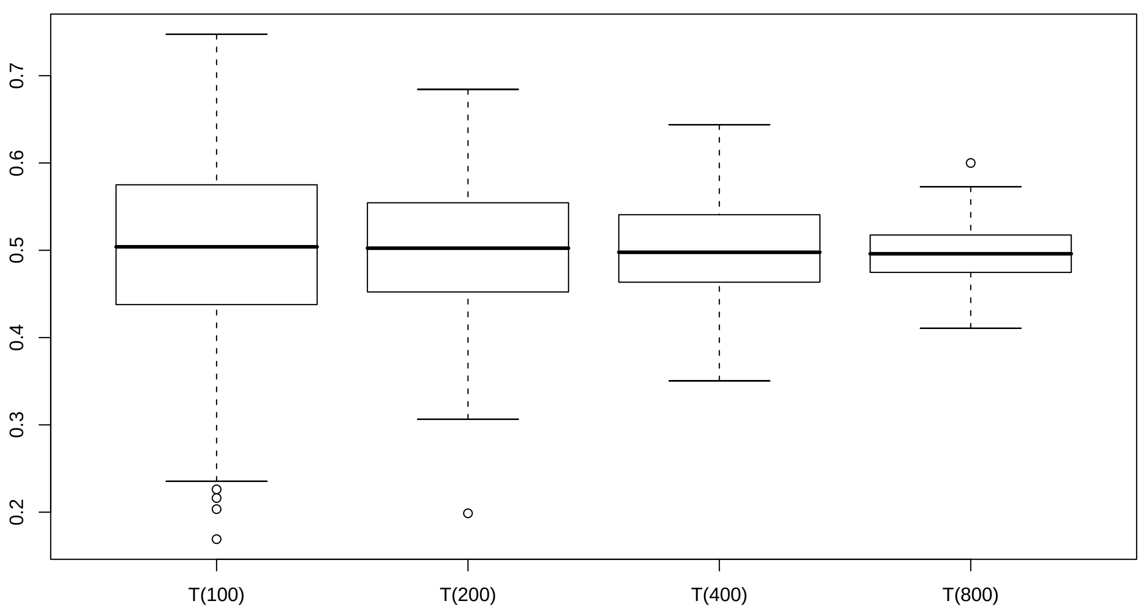

Figure 4. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.5$, $p_{22} = 0.5$, $\phi_1 = 0.9$, $\theta_1 = 0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

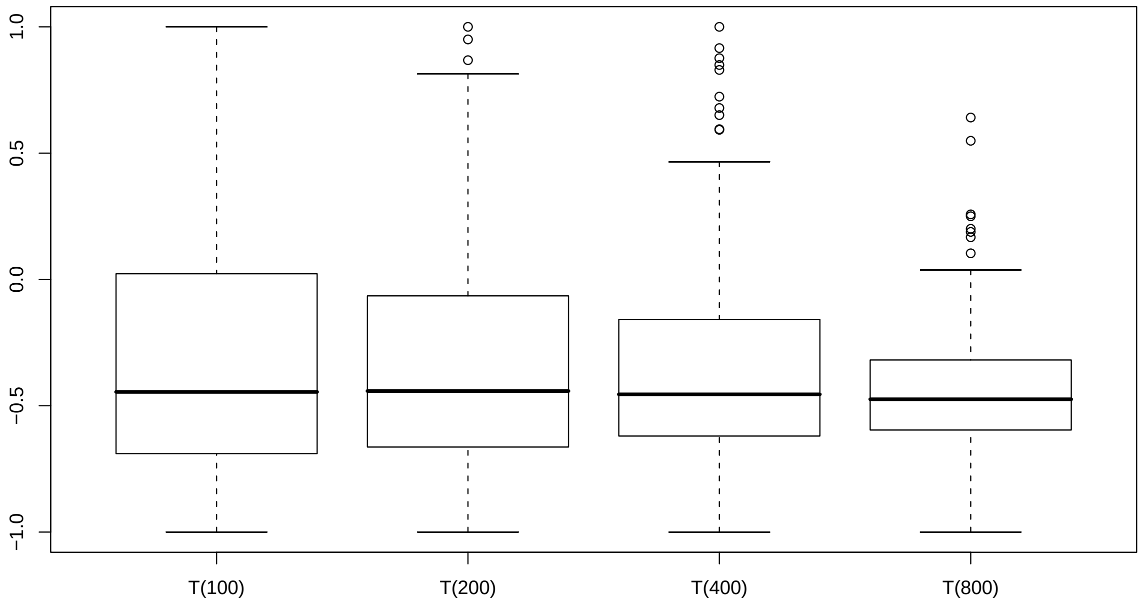

Figure 5. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.95$, $p_{22} = 0.95$, $\phi_1 = 0.6$, $\theta_1 = -0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

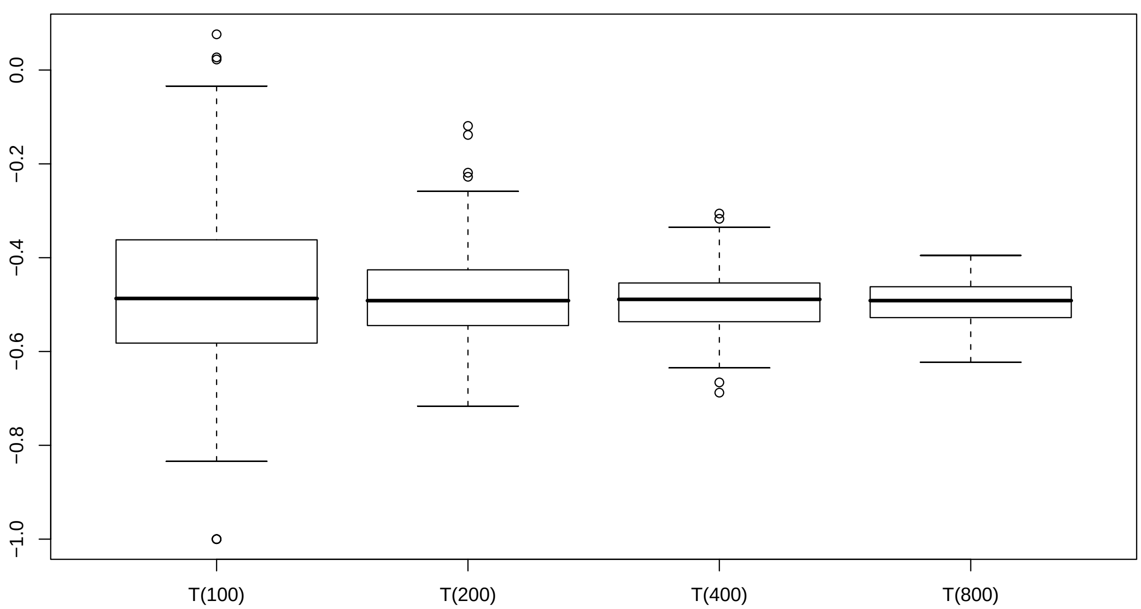

Figure 6. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.95$, $p_{22} = 0.95$, $\phi_1 = 0.9$, $\theta_1 = -0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

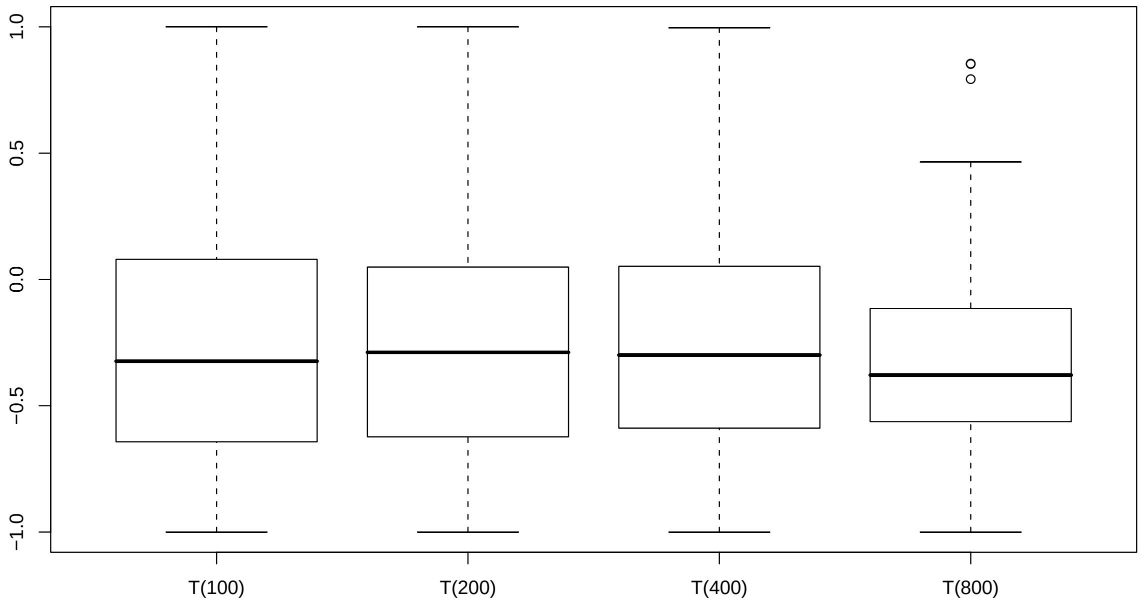

Figure 7. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.5$, $p_{22} = 0.5$, $\phi_1 = 0.6$, $\theta_1 = -0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

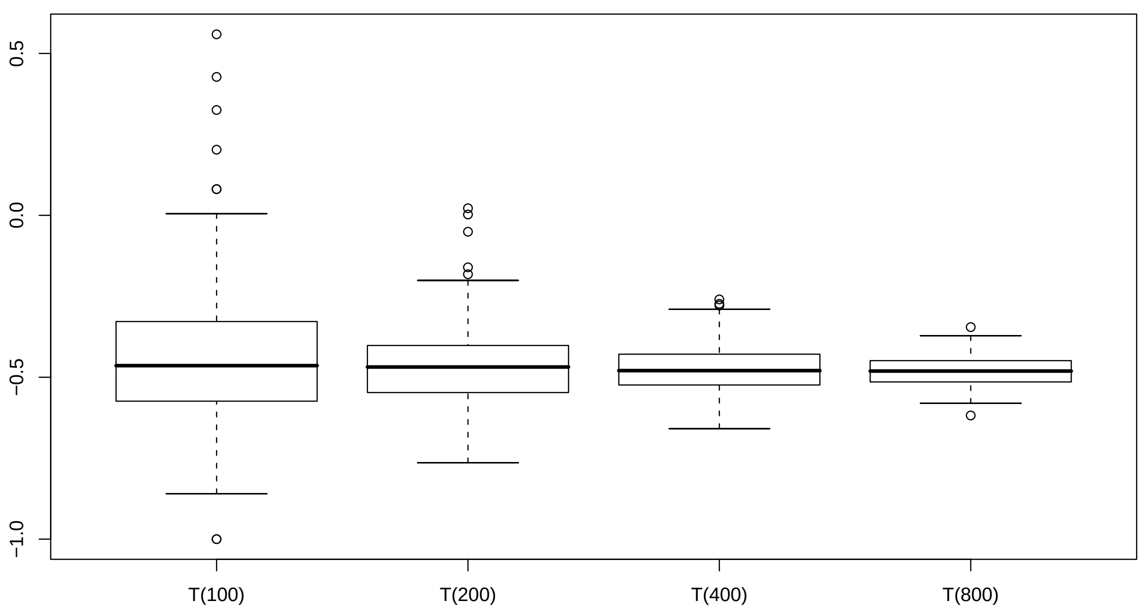

Figure 8. Box-plots of the estimated θ_1 from the model defined in (18) with 200 realizations. The parameters are set as $p_{11} = 0.5$, $p_{22} = 0.5$, $\phi_1 = 0.9$, $\theta_1 = -0.5$, $\mu_1 = 1$, $\mu_2 = 5$, $\sigma_1^2 = 1$, and $\sigma_2^2 = 1.5$.

Number	Author(s)	Title	Date
07-A009	Chao-Chun Chen Wen-Jen Tsay	Estimating Markov-Switching ARMA Models with Extended Algorithms of Hamilton	10/07
07-A008	鮑世亨 黃登興	產品異質性下之市場規模與廠商的群聚	09/07
07-A007	Been-Lon Chen Mei Hsu Chia-Hui Lu	Inflation and Growth: Impatience and a Qualitative Equivalence	06/07
07-A006	Chia-Chang Chuang Chung-Ming Kuan Hsin-Yi Lin	Causality in Quantiles and Dynamic Stock Return-Volume Relations	06/07
07-A005	Shu-Chun Susan Yang	Do Capital Income Tax Cuts Trickle Down ?	05/07
07-A004	Wen-Jen Tsay	The Fertility of Second-Generation Political Immigrants in Taiwan	04/07
07-A003	Wen-Jen Tsay	Estimating Long Memory Time-Series-Cross-Section Data	04/07
07-A002	Meng-Yu Liang C.C. Yang	On the Budget-Constrained IRS : Equilibrium and Efficiency	01/07
07-A001	Kamhon Kan Yen-Ling Lin	The Labor Market Effects of National Health Insurance: Evidence From Taiwan	01/07
06-A015	Chung-cheng Lin C.C. Yang	Reciprocity and Downward Wage Rigidity	12/06
06-A014	Chung-cheng Lin C.C. Yang	The Firm as a Community Explaining Asymmetric Behavior and Downward Rigidity of Wages	12/06
06-A013	林忠正 黃璀娟	補習是一種社會風俗	11/06
06-A012	Pei-Chou Lin Deng-Shing Huang	Technological Regimes and Firm Survival: Evidence across Sectors and over Time	11/06
06-A011	Deng-Shing Huang Yo-Yi Huang Cheng-Te Lee	Technology Advantage and Trade : Home Market Effects Revisited	10/06
06-A010	Chung-Ming Kuan	Artifcial Neural Networks	09/06
06-A009	Wei-Ming Lee Chung-Ming Kuan	Testing Over-Identifying Restrictions without Consistent Estimation of the Asymptotic Covariance Matrix	09/06
06-A008	Chung-Ming Kuan. Yu-Wei Hsieh	Improved HAC Covariance Matrix Estimation Based on Forecast Errors	09/06
06-A007	Yu-Chin Hsu Chung-Ming Kuan	Change-Point Estimation of Nonstationary *I(d)* Processes	09/06

06-A006	Yuko Kinishita Chia-Hui Lu	On the Role of Absorptive Capacity: FDI Matters to Growth	08/06
06-A005	Kamhon Kan	Residential Mobility and Social Capital	07/06
06-A004	Kamhon Kan	Cigarette Smoking and Self –Control	07/06
06-A003	林忠正	懲罰怠惰、流失人才	06/06
06-A002	Shin-Kun Peng Takatoshi Tabuchi	Spatial Competition in Variety and Number of Stores	02/06
06-A001	Mamoru Kaneko J. Jude Kline	Inductive Game Theory: A Basic Scenario	01/06
05-A011	Amy R. Hwang	An Ecological-Economic Integrated General Equilibrium Model	12/05
05-A010	Juin-jen Chang Ming-fu Shaw Ching-chong Lai	A “Managerial” Trade Union and Economic Growth	12/05
05-A009	Lin-Ti Tan	Spatial Economic Theory of Pollution Control under Stochastic Emissions	10/05
05-A008	Kamhon KAN Wei-Der TSAI	Entrepreneurship and Risk Aversion	10/05
05-A007	江豐富	台灣縣市失業率的長期追蹤研究—1987- 2001	08/05
05-A006	Shin-Kun Peng, Jacques-Francois Thisse Ping Wang	Economic Integration and Agglomeration in a Middle Product Economy	06/05
05-A005	譚令蒂 洪乙禎 謝啓瑞	論藥價差	05/05
05-A004	Lin-Ti Tan Yan-Shu Lin	Spatial Monopoly Pricing in a Stochastic Environment	05/05
05-A003	Been-Lon Chen Shun-Fa Lee	Congestible Public Goods and Indeterminacy in a Two-sector Endogenous Growth Model*	03/05
05-A002	C. Y. Cyrus Chu Hung-Ken Chien	The Optimal Decoupled Liabilities: A General Analysis	02/05
05-A001	Cyrus C.Y. Chu Hung-Ken Chien	Durable-Goods Monopolists, Network Effects and Penetration Pricing	02/05
04-A015	Been-Lon Chen	Multiple Equilibria in a Growth Model with Habit Persistence	11/04
04-A014	C. Y. Cyrus Chu R. R. Yu Ruey S. Tsay	A New Model for Family Resource Allocation Among Siblings: Competition, Forbearance, and Support	05/04