# A Complete Procedure For Estimating Hidden Markov Models with Application in Locating Structural Breaks

Christos Ntantamis [*]

CREATES, University of Aarhus

October 13, 2008

**Abstract**

Testing for structural breaks and identifying their location is essential for econometric modeling. In this paper, a Hidden Markov Model (HMM) approach is used in order to perform these tasks. Breaks are defined as the data points where the underlying Markov Chain switches from one state to another. The estimation of the HMM is conducted using a variant of the Iterative Conditional Expectation-Generalized Mixture (ICE-GEMI) algorithm proposed by Delignon et al. (1997), that permits analysis of the conditional distributions of economic data and allows for different functional forms across regimes. The locations of the breaks are subsequently obtained by assigning states to data points according to the Maximum Posterior Mode (MPM) algorithm. The Integrated Classification Likelihood -Bayesian Information Criterion (ICL-BIC) allows for the determination of the number of regimes by taking into account the classification of the data points to their corresponding regimes. The performance of the overall procedure, denoted IMI by the initials of the component algorithms, is validated by two sets of simulations; one in which only the parameters are permitted to differ across regimes, and one that also permits differences in the functional forms. The IMI method performs well in both sets. Moreover, when it is compared to the Bai and Perron (1998) method, which is plausible for the first set of simulation, its performance is superior in the assessing the number of breaks and their respective locations.

**Keywords:** Structural change; Hidden Markov Model; Regime Swithcing; Bayesian Segmentation.
**JEL Classification Numbers:** C13, C22, C52

---

[*] Address: School of Economics & Management, University of Aarhus, Building 1322, 8000 Aarhus C, Denmark, telephone: +45 89422140, email: *cntantamis@creates.au.dk*. The author is grateful to John W. Galbraith for his helpful comments. Center for Research in Econometric Analysis of Time Series, CREATES, is funded by The Danish National Research Foundation.

# 1 Introduction

In pursuing quantitative research, econometricians are frequently able to justify on theoretical grounds a choice for the functional form of an economic model. Rarely, however, are they able to justify a maintained hypothesis that the coefficients of a model should be stable over a sample interval. Thus the stability of estimated coefficients is an empirical question that should not be ignored. Otherwise, the consequences can be rather severe. Quoting Hansen (2001):

> "Structural change is pervasive in economic time series relationships, and it can be quite perilous to ignore. Inferences about economic relationships can go astray, forecasts can be inaccurate, and policy recommendations can be misleading or worse."

For example, the assumptions of the standard linear model, such as the parameter constancy, are unrealistic in many economic applications. Models that do not account sufficiently for structural change are misspecified and inferences may then suggest excessive persistence (Perron 1989, Andrews and Zivot 1992, and Lumsdaine and Papell 1997). Thus it is important to subject any estimated linear model to various specification tests before the model is used for inference or forecasting purposes. Consequently, there is a great amount of work in the current economic literature that focuses on tests for detecting the existence of structural breaks of unknown timing, and on tests that also allow one to estimate the timing of a structural break.

The classical test for detecting structural change is typically attributed to Chow (1960); the problem with this test is that it can only be used when the location of the break point is known a priori. When the location of the break point(s) is unknown, which is what is encountered in most economic applications ,there exist two main streams of literature. The first stream considers tests that make use of F-type test statistics (Quandt 1960,Andrews 1993, Andrews and Ploberger 1994, and Hansen 1997), whereas the second one uses generalized fluctuation tests, e.g. CUSUM type tests (Brown et al. 1975,Krämer et al. 1988, and Alt,Krämer and Ploberger 1992).

Both streams focus mostly in determining the existence of structural breaks but not their exact locations. An obvious candidate as an estimate for the location of the break is the date that yields the largest value of the Chow (1960) test sequence. Nevertheless, this would be a good estimate only in the case of linear regressions when the Chow test is constructed with the "homoskedastic" form of the covariance matrix. In regression models, an appropriate method to estimate the parameters – including the break point – is least squares. Operationally, the sample is split at each possible breakpoint, the other parameters estimated by ordinary least squares and the sum of squared errors calculated and stored. The least squares break point estimate is the date that minimizes the full-sample sum of squared errors. A theory of least squares estimation has been developed by Jushan Bai and coauthors. Bai (1994, 1997a) derived the asymptotic distribution of the break point estimator and showed how to construct confidence intervals for it. Additionally, Bai and Perron (1998) developed tests for multiple structural changes and suggested procedures for the simultaneous estimation of multiple breakdates. Their method is sequential, starting by testing for a single structural break. It the test rejects the null hypothesis that there is no structural break, the sample is split in two (based on an estimate of the break date) and the test is reapplied to each subsample. This procedure continues until each subsample test fails to find evidence of a break. Given the computational complexity of the approach, Bai and Perron (2003) suggested an efficient way of implementing their algorithm. Finally, Perron and Qu

(2006) extended the Bai and Perron (1998) methodology in order to account for restrictions posed on the parameters.

A totally different approach into the problem of structural breaks is the employment of regime switching models, also known as Hidden Markov Models (HHM) (Hamilton 1989, Engel 1994, Schaller and van Norder 1997 and Engel and Hamilton 1990, Marsh 2000 for incorporating exogenous variables and autoregressive terms within the models). This approach considers the locations of the breaks as the points where a switch from one regime to another occurs. Compared with the approaches mentioned above, such breaks are no longer regarded as the outcome of a perfectly foreseeable, deterministic event, but they are instead random variable themselves. The HHM allows to estimate the probability laws that govern switches from one regime to another. Moreover, even if the true underlying model does involve an event happening only once and there is no reason to assume that it will reverse, the HHM still can identify the location of this break but it will fail when forecasting takes place as it will assume a nonnegative probability for a regime switch.

The main features in the current economic literature involving a frequentist estimation of HMM are: a) the number of regimes is given *a priori*, and usually set according to some economic theory to be equal to two, and it is not a result of a statistical procedure, b) the location of regime switches are obtained by the expected durations of each regime, c) the models across regimes are the same and only their parameters are allowed to differ, and d) the estimation is performed using variants of the Expectation Maximization (EM) algorithm. Resorting to Bayesian estimation, some of the potential drawbacks of the frequentist approach can be dealt with. There has been some work on the identification of the location of breaks and of the determination of the number of regimes, either by comparing Bayes' Factors or by modeling the location of a break as a random variable (Pesaran et al. 2004).

Here, an alternative estimation method in a frequentist approach setting will be extended. This algorithm is the Iterative Conditional Estimation (ICE) and was introduced by Pieczynski (1992). This algorithm is no longer based on the notion of likelihood but on that of conditional expectation. Therefore, this approach is of wider application because it encompasses probability distributions that have both a discrete and a continuous part, when the notion of likelihood is no longer valid. The employment of the ICE algorithm in engineering applications, and especially image segmentation, was advocated in a series of papers by Pieczynski and coauthors (Pieczynski 1994, Fjoroft et al. 2003) for the case of i.i.d data. The most common variation of this algorithm is the Iterative Conditional Estimation-Generalized Mixture (ICE-GEMI) algorithm proposed by Pieczynski(1996), which allows data to be generated according to different distributions across regimes.

The main feature in the implementation of the ICE algorithm is its semi-Bayesian nature. It is an iterative algorithm, which at each step simulates a path of the hidden states based on the current estimates of the transition probabilities of the hidden Markov Chain, conditional on the observable data. When the simulated path of states is available, the observed data points are allocated to the different regimes, and they are subsequently used to recover the parameters for the observations' distributions for each regime. If the path is "close" to the actual, and hidden, path, the parameter estimates will be such that a similar path will be generated at the next iteration, and thus lead to the convergence of the algorithm. This artificial division of the data allows for different models, i.e. different observations' distributions, to be considered for each regime as long as the parameters are consistently estimated.

In this paper, I modify the ICE-GEMI algorithm so as to consider cases that are encountered in economics applications. Instead of the unconditional distribution of the observed data series, the distributions of the dependent variable conditional on a set explanatory variables are examined. Moreover, the models across regimes are allowed to differ; the most appropriate model is picked among members of a pre-determined set of models. All models are estimated using the data points that belong to each regime; the one that has the smallest value for the Bayesian Information Criterion is assigned to the particular regime. The main assumption is that all models can be consistently estimated using the data points of each regime. There are no other papers in the economic literature, to the author's knowledge, that allow for such differences of functional forms across regimes and thus this approach provides such flexibility for the first time.

Two more issues are also addressed. The first involves the determination of the location of the regime switching points. The Bayesian segmentation approach is used for determining the optimal sequence of the regimes by assigning data points to regimes according to probabilistic arguments for the likelihood of data point(s) belonging to a particular regime. In this paper, the Maximum Posterior Mode (MPM) approach, which assigns each data point to the regime that is most likely to have generate it, is used for its simplicity as it is directly obtained through the ICE-GEMI estimation.

The second issue addresses the problem of choosing the "optimal" number of regimes that best describe the data. The parameters of Hidden Markov Models are not identified in the statistical sense. Thus, formal testing for the number of regimes is cumbersome and computationally intensive, when it involves Likelihood Ratio tests. The employment of information criteria, which is less demanding, is also used despite the problems that also arise due to lack of identification. Among the information criteria, the Integrated Classification Likelihood-BIC (ICL-BIC) criterion has a number of advantages. It takes into account the classification of the data points to the different regimes in order to penalize the lack of fit, and it is also based on the computation of the integrated likelihood, which is the basis of the BIC. Moreover, it can be easily calculated; it is basically the BIC plus a term that is obtained as soon as the model estimation is completed. Therefore, it will be used in this paper in order to determine the number of regimes in the HMM.

Overall, this paper introduces into the economic literature a complete procedure for estimating all the aspects of a Hidden Markov Model. Regarding model estimation, for a given number of regimes, a variant of the ICE-GEMI algorithm is used, which allows model heterogeneity among regimes. The locations of regime switches are obtained by the MPM algorithm that assigns points according to probability of occurrence, and the "optimal" number of regimes is derived according to the ICL-BIC criterion. Henceforth, this computational procedure will be denoted as IMI by the initials of its three components.

Simulations are used in order to assess the ability of the proposed procedure to pick the correct number of regimes, to identify the location of regime changes, and to estimate the parameters of the models for each of the regimes. There are two sets of specifications. The first involves data generating processes in the form of a multiple linear regression model whose parameters are allowed to differ. The results are compared with results obtained using the Bai and Perron (1998) method. The second specification involves cases where models are allowed to differ across regimes. As mentioned previously, no other methodology exists for comparison purposes.

The remaining of the paper is organized as follows. Section 2 discusses the details of the IMI procedure; introduces the ICE-GEMI algorithm, explains the MPM approach and provides an overview of the task of estimating the number of regimes in the Hidden Markov Models. The description of the simulation specifications and the corresponding results are demonstrated in Section 3, and Section 4 concludes the chapter.

# 2 Methodology

## 2.1 A Brief review of Hidden Markov Models

In what follows, only a brief description of the Hidden Markov Models is provided; for a more detailed review refer to Bilmes (2006), or Rabiner (1989).

A Hidden Markov Model (HMM) is a discrete -time stochastic process including an underlying finite-state Markov Chain (state sequence) and a sequence of random variables whose distributions depend on the state sequence of random variables whose distributions depend on the state sequence only through the current state (observation sequence). The state sequence is not observable, and hence all the conclusions about the process must be made using only the observation sequence[1].

A more formal definition, which presents the probabilistic relationships between the state sequence and the observation sequence, is [2]:

*Definition*: A Hidden Markov Model (HMM) is a collection or random variables consisting of a set of $T$ discrete scalar variables $X_{1:T}$ and a set of T other variables $Y_{1:T}$ which may be either discrete or continuous (and either scalar - or vector-valued). These variables, collectively, possess the following conditional independence properties:

$$\{X_{t:T}, Y_{t:T}\} \perp \{X_{1:t-2}, Y_{1:t-1}\} \mid X_{t-1} \tag{1}$$

and

$$Y_t \perp \{X_{\neg t}, Y_{\neg t}\} \mid X_t \tag{2}$$

for each $t \in 1 : T$. No other conditional independence properties are true in general, unless they follow from (1) and (2).

An HMM will be, given the definition, any joint probability distribution over an appropriately typed set of random variables $(X, Y)$ that obeys the stated set of conditional independence rules. The two conditional independence properties imply that, for a given $T$, the joint distribution over all variables may be expanded as follows (using the Chain rule and the definition equations):

$$P\left(y_{1:T}\right) = P\left(y_1\right) \prod_{t=2}^{T} P\left(x_t \mid x_{t-1}\right) \prod_{t=1}^{T} P\left(y_t \mid x_t\right) \tag{3}$$

Thus, in order to parameterize an HMM, one needs the following quantities:

(i) The distribution over the initial state variable: $P\left(x_1\right)$.

(ii) The conditional transition distributions for the first-order Markov Chain: $P\left(x_t \mid x_{t-1}\right)$.

(iii) The conditional distribution for the other variables: $P\left(y_t \mid x_t\right)$.

It can be seen that these quantities correspond to the classic HMM definition, where $X_t$ is the hidden Markov Chain and $Y_t$ are the observed data. Specifically, the initial (not necessarily stationary, see Resnick 1992 for details) distribution is labeled $\pi$ which is a vector of length equal to the number of the different states.

---

[1] Henceforth, the terms state and regime will be used interchangeably.

[2] The following notation will be employed:$X_{s:q}, s < q = \{X_s, X_{s+1}, \ldots, X_q\}, X_{<s} \triangleq \{X_1, X_2, \ldots, X_{s-1}\}$, and $X_{\neg t} \triangleq X_{1:T} \setminus X_t = \{X_1, X_2, \ldots, X_{t-1}, X_{t+1}, X_{t+2}, \ldots, X_T\}$.

Then, $\pi_i = P(X_1 = i)$, where $\pi_i$ is the $i^{th}$ element of $\pi$. The observation probability distributions are denoted $b_j(y_t) = P(Y_t = y_t \mid X_t = j)$ and the associated parameters depend on the $b_j(y)$'s family of distributions. Furthermore, the Markov Chain is typically assumed to be time-homogeneous, with stochastic transition matrix $\mathbf{A}$, where $A_{ij} = P(X_t = j \mid X_{t-1} = i)$, for all $t$. HMM parameters are often symbolized collectively as $\lambda \triangleq (\pi, \mathbf{A}, \mathbf{B})$, where $\mathbf{B}$ represents the parameters corresponding to all the observation distributions.

## 2.2 The ICE-GEMI algorithm

### 2.2.1 Introduction

The HMM parameters, $\lambda \triangleq (\pi, \mathbf{A}, \mathbf{B})$, are usually estimated with an algorithm advocated in a series of papers by Baum and co-authors (e.g. Baum et al 1970), which is known as the Baum-Welsh algorithm. This algorithm belongs to the Expectation Maximization (EM) algorithm family which was formally introduced by Dempster et al (1977), in order to feasibly estimate model parameters using maximum likelihood. The core of the Baum-Welsh algorithm is the recursive estimation of the forward and backward probabilities ($\alpha$ and $\beta$ respectively). When these are recovered and with further assumptions regarding the form of the observations' distributions, the recursive estimation of the HMM parameters is straightforward.

In this paper, focus will be put on an alternative algorithm: the Iterative Conditional Estimation (ICE) that was introduced by Pieczynski (1992). This algorithm is based on the notion of conditional expectation, and thus it encompasses probability distributions that have both a discrete and a continuous part, which is a case where the notion of likelihood is no longer valid [3]. The employment of the ICE algorithm in engineering applications, and especially image segmentation, was advocated in a series of papers by Pieczynski and coauthors (Pieczynski 1994, Fjoroft et al. 2003 ). The main feature in the implementation of the ICE algorithm is the simulation of the hidden states based on the current estimates of the transition probabilities of the hidden Markov Chain conditional on the observable data. When a simulated path of states is available, the observed data points are allocated to the different regimes. This artificial division of the data allows for different models, i.e. different observations' distributions, to be considered for each regime as long as the models' parameters are consistently estimated.

The most common variation of this algorithm, that takes advantage of the aforementioned flexibility, is the Iterative Conditional Estimation-Generalized Mixture (ICE-GEMI) algorithm proposed by Pieczynski(1996). The feature of the generalized mixture estimation is that even though the exact form of the observation distributions $b_j$ is unknown, they are assumed to belong to a particular given set of distributional forms. Thus, it will be considered that $b_j \in F = \{F_1, F_2, \ldots, F_M\}$, where $F_i$ represents a family of distributions, and $b_j$ is the distribution of the observations $(Y_t)$ given the Hidden Markov Chain $(X_t)$ is in state $j$ (out of $n$ possible states). Consequently, for each $\{b_j\}_{j=1,\ldots,n}$ one must determine: a) the family that it belongs to, according to some decision rule, and b) the values of the corresponding parameters.

### 2.2.2 Description of the ICE-GEMI algorithm for the i.i.d. case

Since the ICE-GEMI has not been used in the economics' literature before, it is of interest to provide a more detailed description of it. Thus, in the current section, a more formal description for the ICE-GEMI

---

[3] Delmas (1997) derived the exact conditions for the equivalence of the EM and the ICE algorithms for the case of the exponential family of distributions.

algorithm is provided for the case that the observed data are comprised of a single random variable and i.i.d., as it is in the case of image segmentation applications. It was already mentioned that the ICE-GEMI algorithm is an iterative procedure. Thus, there are a number of assumptions that need to be satisfied before its implementation in order to ensure convergence. In particular, the following assumptions must hold for the case of $(X_t, Y_t)$ being i.i.d.:

1. An estimator of the parameter vector $\zeta$ from $X_{1:T}$ exists: $(\hat{\zeta} = \hat{\zeta}\left(X_{1:T}\right))$. These parameters are used for simulating the state paths.
2. We can simulate $X_{1:T}$ according to its distribution conditional on $Y_{1:T}$.
3. Each observation distribution $b_j$ is characterized by a parameter vector $\theta^j$.
4. There are M estimators $\left\{\hat{\theta}^1, \hat{\theta}^2, \ldots, \hat{\theta}^M\right\}$ so that for a sample $\mathbf{w} = \{w_1, \ldots, w_T\}$ generated by the distribution $b_j\left(x \mid \theta_i\right) \in F_i$, $\hat{\theta}^i = \hat{\theta}^i(\mathbf{w})$ estimates $\theta_i$.
5. A decision rule $D$ exists, such that, for any sample $\mathbf{w} = \{w_1, \ldots, w_T\}$, and any $\{f_1, f_2, \ldots, f_M\} \in F_1 \times F_2 \times \cdots \times F_M$, the rule $D$ will associate the sample $\mathbf{w}$ with the "closer" density among $\{f_1, f_2, \ldots, f_M\}$ according to some criterion.

If the aforementioned assumptions hold, the ICE-GEMI algorithm will be (at step q):

- current prior parameters: $\zeta^q$
- current observation distributions: $b_j = f_j^q, \quad j = 1, \ldots, n$
- iterative procedure:
  - (i) Simulate $x^q$ to be a realization of $X_{1:T}$ according to $\zeta^q$, $\{f_j^q\}_{j=1,\ldots,n}$, and conditional on $y_{1:T}$.
  - (ii) Calculate $\zeta^{q+1} = E_q\left[\hat{\zeta}\left(X_{1:T}\right) \mid y_{1:T}\right]$. If this is not feasible, take $\zeta^{q+1} = \hat{\zeta}\left(x_{1:T}^q\right)$.
  - (iii) For each $i = 1, \ldots, n$, consider the set of events $S_i^q = \{s \in [1, 2, \ldots, T] \mid x_s^q = i\}$ and the corresponding subsets of $y_{1:T}$, $y_i^q = \left(y_s\right)_{s \in S_i^q}$, and using these subsets estimate the M parameters: $\theta_i^1 = \hat{\theta}^1\left(y_i^q\right), \ldots, \theta_i^M = \hat{\theta}^M\left(y_i^q\right)$.
  - (iv) For each $i = 1, \ldots, n$, consider the decision rule: $D\left(y_i^q\right) \in \left\{f_{\theta_i^1}, \ldots, f_{\theta_i^M}\right\}$.
  - (v) Update the observation distributions: $b_j = f_j^{q+1} = D\left(y_j^q\right), \quad j = 1, \ldots, n$.

In order to increase the accuracy of the results, we can increase the number of realizations of $X_{1:T}$, which are obtained in step (i) of the iterative procedure, to W, i.e. $x_1^q, x_2^q, \ldots, x_W^q$.

Then, the updated estimate for the parameter $\zeta^{q+1}$ will be given by:

$$\zeta^{q+1} = \frac{1}{W} \sum_{w=1}^{W} \hat{\zeta}\left(x_w^q\right) \tag{4}$$

In practice, a single realization will suffice[4].

---

[4] According to Delignon et al. 1997, even a single realization is adequate in image segmentation applications. Nevertheless, they do not provide any specific conditions for their remark.

Prior to the implementation of the ICE-GEMI algorithm a verification of the validity of the assumptions is required. This is straightforward when $(X_t, Y_t)$ being iid.

1. The parameter $\zeta$ will be $\zeta = \{c_{ij}\}_{i=1,\ldots,n}^{j=1,\ldots,n}$, where:

$$c_{ij} = P\left(X_t = i, X_{t+1} = j\right) \tag{5}$$

   and it does not depend on the time $t$. These quantities allow the derivation of the initial state distribution and the state-transition probability matrix:

$$\pi_i = \textstyle\sum_j c_{ij} \quad A_{ij} = \frac{c_{ij}}{\sum_j c_{ij}} \tag{6}$$

   The estimator of $\hat{c}_{ij}$ can be based on the empirical frequencies, as obtained in the simulated path:

$$\hat{c}_{ij} = \frac{1}{T-1} \sum_{t=1}^{T-1} 1_{[x_t=i, x_{t+1}=j]} \tag{7}$$

2. The hidden Markov Chain $X_{1:T}$ is, conditionally on $y_{1:T}$, a non-stationary Markov Chain. The transition matrix at time $t$ is given by:

$$A_{ij}^t = \frac{\xi_t(i,j)}{\gamma_t(i)} \tag{8}$$

   where:

$$\gamma_t(i) = P\left(X_t = i | y_{1:T}\right) \tag{9}$$

   and

$$\xi_t(i,j) = P\left(X_t = i, X_{t+1} = j | y_{1:T}\right) \tag{10}$$

   The quantities $\gamma_t(i), \xi_t(i,j)$ can be determined using the forward $(\alpha)$ and backward$(\beta)$ probabilities according to the Baum-Welch algorithm [5]:

$$\xi_t\left(i,j\right) = \frac{\alpha_t\left(i\right) A_{ij} f_j\left(y_{y+1}\right) \beta_{t+1}\left(j\right)}{\sum\limits_{l=1}^{n} \sum\limits_{m=1}^{n} \alpha_t\left(l\right) A_{lm} f_m\left(y_{y+1}\right) \beta_{t+1}\left(m\right)} \tag{11}$$

   and

$$\gamma_t\left(i\right) = \frac{\alpha_t\left(i\right) \beta_t\left(i\right)}{\sum\limits_{l=1}^{n} \alpha_t\left(l\right) \beta_t\left(l\right)} \tag{12}$$

   Hence, an a posteriori realization of $X_{1:T}$ can be simulated:

   - The state $X_1 = i$ of the first element of the chain is drawn randomly according to the marginal *a posteriori* distribution $\gamma_1(i)$.

---

[5] The formulae for the the calculation of the forward and backward probabilities can be found in Bilmes(2006). In this paper, in order to address the numerical instability issues that appear in their calculations, the Devijer et al. (1988) approach, which replaces the joint probabilities with *a posteriori* probabilities, was employed.

- For each new element $t = 1, \dots, T-1$, the transition probabilities $A_{ij}^t$ are given, the state of the previous element $X_t = i$ is kept fixed, and $X_{t+1}$ is obtained by random sampling according to this distribution.

At this point, two notes must be made. First, the GEMI nature of this algorithm involves the choice of the distributional form for the observations' distributions among a predefined set of candidates. Thus, when the sample is split according to the simulated path, estimation of the parameters for each of the candidate distributional forms takes place and subsequently a decision rule determines which of those forms is the most appropriate. A simpler version will be the use of a single distributional form, e.g. Normal. In such a case, only the parameter estimation takes place. The second note involves an intuitive explanation of the mechanism of the algorithm. Given a simulated path, we obtain estimates of the parameters of the observed data distribution under each regime. If we focus on the probability of occurrence of a single data point, then it is expected that a higher probability will be produced under the model corresponding to the correct regime, i.e. the model that served as the DGP for the data points belonging to the particular regime, compared to the wrong regime. This higher probability is then fed into the algorithm in order to produce a higher probability of occurrence for the particular regime (as expressed by $A^t(i,j)$) when the regime path is simulated. This procedure continues until convergence. Thus, a good split of the data will result in better estimates of the distributions' parameters for each regime, which in turn will yield higher probabilities of regime occurrence for these points, which will finally lead to an even better split of the data.

### 2.2.3 Modifying the ICE-GEMI for including explanatory variables

The ICE-GEMI algorithm as presented in the previous section can only be applied to cases where the interest lies in the unconditional distribution of the observed data $Y_t$. This is not an issue when image segmentation applications are considered but it poses a lot of restrictions in economic data applications. In such cases, the availability of extra information in the form of explanatory variables requires us to consider the conditional distribution of the observed data. Regardless, the ICE algorithm can be easily adapted to these cases as follows.

Consider the observed data $y_t$ which, under regime $j$, are generated according to:

$$y_t = g_j\left(Z_t; \theta_j\right) + \epsilon_t \tag{13}$$

where $Z$ is a vector of explanatory variables, and $g_j$ is a function, whose form can be allowed to differ across regimes and:

$$\epsilon_t \sim i.i.df\left(0, \sigma_j^2\right)$$

The focus will now be on the residuals from the model estimation. At the end of the previous section, it was mentioned that ICE algorithm converges as higher probabilities that data points have been assigned "correctly" to regimes lead to higher probabilities of repeating this assignment through the simulated path. How will this work in the case of the model in (13)?

The following changes need to be made to the ICE-GEMI algorithm presented in the previous section.

1. It is assumed that the functional form $g_j\left(Z_t; \theta_j\right)$ may belong to a predetermined set of M different models, e.g. linear regression, AR(1).

2. For each $i = 1, \ldots, n$, consider the set of events $S_i^q = \{s \in [1, 2, \ldots, T] \mid x_s^q = i\}$ and the corresponding subsets of $y_{1:T}$, $y_i^q = (y_s)_{s \in S_i^q}$,$Z_{1:T}$, $Z_i^q = (Z_s)_{s \in S_i^q}$, and using these subsets estimate the M parameters for each of the candidate models:

$$\begin{aligned}\left(\theta_i^1, \sigma_i^1\right) = \left(\hat{\theta}^1\left(y_i^q, Z_i^q\right), \hat{\sigma}^1\left(y_i^q, Z_i^q\right)\right), \ldots, \left(\hat{\theta}^M\left(y_i^q, Z_i^q\right), \hat{\sigma}^M\left(y_i^q, Z_i^q\right)\right) = \\ = \left(\hat{\theta}^M\left(y_i^q, Z_i^q\right), \hat{\sigma}^M\left(y_i^q, Z_i^q\right)\right)\end{aligned} \tag{14}$$

3. For each regime $i = 1, \ldots, n$, consider the decision rule:

$$D\left(y_i^q, Z_i^q\right) \in \left\{g_{\theta_i^1}, \ldots, g_{\theta_i^M}\right\} = g_i^*(\hat{\theta}_i) \tag{15}$$

4. Update the observation distributions $b_j$ using the generated residuals:

$$\hat{\epsilon}_t^j = y_t - g_i^*\left(Z_t; \hat{\theta}_j\right) \tag{16}$$

   which results in: $b_j = f\left(\hat{\epsilon}^j; 0, \hat{\sigma}_j\right), \quad j = 1, \ldots, n.$

As in the case of ICE-GEMI, the mechanism is obvious: at the time points that correspond to the "correct" regime and the correct model the residuals will be small and thus they will yield large probabilities in the observations' distributions. One of the advantages of this approach is the availability, under each regime, of more types of models that may describe the data in a better fashion. This feature of the algorithm is especially important when the number of regimes needs to be estimated, as it will become apparent is a subsequent section.

One of the issues that requires some attention is the choice of the decision rule for assigning to each regime the most "suitable" model. In the original ICE-GEMI framework, Pieczynski(1996) suggested using a Kolmogorov-Smirnov distance measure between the theoretical values for the cumulative density function, obtained by the proposed distributional forms and the empirical values obtained by the available data points. A direct analogue in the extended setting would have been the comparison between the theoretical values obtained by the models and the values that are obtained by using Kernel regression. Even though this approach is appealing, it adds computational burden to the model[6].

Instead, I will use the Bayesian Information Criterion (BIC) as the decision rule[7]. When the model parameters are estimated, the BIC is also estimated for each of the models. The decision rule will assign the model that has the smallest BIC value to the regime under consideration.

## 2.3 Identifying the hidden states

Another part of the analysis will be to uncover the hidden part of the model, i.e. to find the "correct" state sequence. The locations of the switches between regimes can then be considered as the location of structural changes for the parameters of the models. It should be clear that for all but the case of degenerate models, there is no "correct" state sequence to be found. Hence, for practical situations, an optimality criterion

---

[6] Kernel regression is known to require considerable computing time even for a small number of regressors.

[7] Alternative Information Criteria could be used, for example the Akaike Information Criterion and the Hannan-Quin Criterion.

is usually employed to solve this problem in an acceptable way (there are several reasonable optimality criteria that can be imposed). So far in the economic literature, Markov regime switching models identify the location of the switch (break) by the expected durations of the regimes. However, there are better ways to identify them if we use what is known as Bayesian Segmentation. This approach assigns data points to regimes according to probabilistic arguments for the likelihood of data point(s) belonging to a particular regime.

One potential way to do this is the Maximum Posterior Mode (MPM) approach. This approach reconstructs the state sequence by allocating at each time point the state that it is more probable, given the observations and the model, i.e.

$$\hat{x}_t = \arg \max_{i=1,\ldots,n} \gamma_t(i), \qquad 1 \leq t \leq T \tag{17}$$

Although this criterion maximizes the expected number of correct states, there could be some problems with the resulting state sequence (the "optimal" state sequence may, in fact, not even be a valid state sequence). This occurs because MPM determines the most likely state at every instant, without regard to the probability of sequences of states. In order to circumvent this problem, one may attempt to maximize the expected numbers of correct pairs of states $(x_t, x_{t+1})$ or triples of states $(x_t, x_{t+1}, x_{t+2})$. The extreme version of this approach, which is the Maximum A Posteriori (MAP) sequence, is to find the single best state sequence (path), i.e. to maximize $P(X \mid \mathbf{y}, \lambda)$ that is equivalent to maximize $P(X, \mathbf{y} \mid \lambda)$. This is accomplished by the Viterbi algorithm.

Nevertheless, the quantity $\gamma_t(i)$ is calculated at each iteration step, and thus it is readily available. Therefore, for reasons of simplicity, the MPM criterion will be employed.

## 2.4 Detecting the number of regimes in mixture models

The presentation of the HMM estimation using the ICE methodology and the restoration of the hidden state sequence in the previous sections assumes that the number of the different regimes is known *a priori*. This is not the case in the economic literature. Unless there are theoretical grounds (which is rare), the researcher is in the dark in terms of the appropriate number of regimes. In most of the work on inference on the number of states in HMMs, Bayesian or otherwise, the main approach has been to separate the problem of testing for the number of regimes $n$, from the fitting of the mixture model, and hence estimation, for a fixed value $n$. There are two main estimation procedures for obtaining the number of regimes, which are built mainly by considering the likelihood of the model estimated for a pre-determined number of regimes.

One procedure is based on a penalized form of the likelihood. As the likelihood increases with the addition of an extra regime, the likelihood (usually, the log likelihood) is penalized by the subtraction of a term that "penalizes" the model for the number of parameters in it. This leads to a penalized log likelihood, yielding the so called information criteria for the choice of $n$. There exist three main categories of penalized likelihood criteria: a) the information criteria that are obtained as an attempt to minimize the Kullback-Leibler information, such as the Akaike Information Criterion (AIC) proposed by Akaike (1974), b) the information criteria that have been derived in a Bayesian framework but can be applied also in a non-Bayesian framework, such as the Bayesian Information Criterion (BIC) proposed by Schwarz (1978), and c) the information criteria that are based on the classification likelihood, i.e. the likelihood that takes also into account the classification of the data points to each regime, such as the Integrated Classification Likelihood (ICL) criterion proposed by Biernacki et al. (1998).

The other main procedure for deciding on the order of a mixture model is to carry out a hypothesis test, using the likelihood ratio as the test statistic (LR). The LR requires the employment of bootstrapping in order to obtain an assessment of the p-value [8]. This task is computationally demanding and thus less desirable. On the other hand compared to the information criteria, quantities that quantify the confidence in the result, such as a p-value, are produced. Nevertheless, in this paper, a version of a penalized log likelihood criterion (ICL) will be preferred because of its lower burden. The proposed criterion takes into account the particular nature of the Hidden Markov Model, and it has been demonstrated by simulations to perform well in the context of mixture models. In the subsequent sections, a more detailed discussion of the ICL criterion will be provided.

Another practical issue arises with the parametric specification for the models under each regime. Even though the assumed models are being taken to reflect the number of regimes in the sample data, there may be cases that spurious regimes are revealed in an attempt to capture aspects of the data that are not properly represented by the assumed models. Thus, it is crucial either to pre-specify flexible models that may capture most of the data characteristics or to allow a choice among different models so as to capture these characteristics. For example, when a normal mixture model is being used to detect the presence of grouping in some data, then there may not be a one-to-one correspondence between the mixture components and the groups if the data have a skewed distribution within some of the groups. This could be solved either by assuming that the different groups in the data are generated according to mixtures of normals, or by allowing the different groups to have been generated by a member of a set of distributions that may include the normal or any type of asymmetric distribution.

### 2.4.1 The Integrated Classification Likelihood Criterion

The last category of information criteria is the ones that take into account the classification of the data points to each regime. The notion of complete-data likelihood $L_c(\Psi)$ is clearer within the Expectation Maximization (EM) framework (Dempster et al. 1977). In this framework, the original data set is augmented by a set of indices $z_{it}$ that represent the classification of each data point $y_t$ to the particular regime $i$ under which the data point was generated and which are called the component indicator variables. Given this, the complete-data likelihood is also called the classification likelihood in the classification context. It is noted by Hathaway (1986) that the HMM log likelihood can be written in terms of the classification likelihood as:

$$\log L(\Psi) = \log L_c(\Psi) - \log k(\Psi) \tag{18}$$

where $\log k(\Psi)$ is the conditional density of the component-indicator variables, given the observed data

$$\log k(\Psi) = \sum_{i=1}^{n} \sum_{t=1}^{T} z_{it} \log \gamma_t(i)$$

---

[8] Unfortunately with mixture models, and their extension the Hidden Markov Models, regularity conditions do not hold for the test statistic of LRT to have its usual asymptotic distribution of chi-squared.

with conditional mean equal to $-EN(\gamma)$

$$EN(\gamma) = \sum_{i=1}^{n}\sum_{t=1}^{T} \gamma_t(i) \log \gamma_t(i) \tag{19}$$

Thus, the criterion that determines the number of regimes that minimizes the classification likelihood, is called the Classification Likelihood Information Criterion (CLC) becomes:

$$-2logL\left(\hat{\Psi}\right) + 2EN(\hat{\gamma}) \tag{20}$$

where $\hat{\gamma}$ are the values of $\gamma_t(i)$ obtained using the parameter estimates. The CLC has a number of disadvantages. First, it tend to overestimate the correct number of regimes when no restrictions are placed on the mixing proportions (Biernacki, Celeux, and Govaert 1999). Secondly, when the regimes are not well separated, the CLC approach to model fitting leads to severely biased estimates of the parameters (McLachlan and Peel 2000).

In an attempt to overcome the problems of CLC and of BIC, Biernacki et al. (2000) are following the Bayesian approach to the construction of the information criterion and they attempt to minimize the integrated classification likelihood. Assuming a Dirichlet prior over the initial probabilities vector ($\pi$), with all its parameters set equal to $\nu$, and applying a BIC approximation, results in the Information Classification Likelihood Criterion (ICL):

$$-2\log L\left(\hat{\Psi}\right) + 2EN(\hat{\gamma}) + 2T\sum_{i=1}^{n} \hat{\pi}_i \log \hat{\pi}_i + d_1 \log T - 2K\left(T\hat{\pi}_1, \ldots, T\hat{\pi}_n\right) \tag{21}$$

where

$$K\left(T\hat{\pi}_1, \ldots, T\hat{\pi}_n\right) = \sum_{i=1}^{n} \log\Gamma(T\hat{\pi}_i + \nu) - \log\Gamma(T + n\nu) - n\log\Gamma(\nu) + \log\Gamma(n\nu) \tag{22}$$

where $d_1$ is the total number of parameters in the regimes's models (not including the number of the regimes). If further approximations are considered in 22 and terms of order o(1) are neglected, the ICL-BIC criterion for selecting the number of regimes $n$ amounts in minimizing the following quantity

$$-2\log L\left(\hat{\Psi}\right) + 2EN(\hat{\gamma}) + d\log T \tag{23}$$

The ICL-BIC variant is easier to calculate and its performance differs only a little from the original, more accurate version of ICL. Moreover, McLachlan and Peel (2000) verified its ability to correctly assess the number of components within finite mixture specifications. Thus, the ICL-BIC Criterion is used in this paper in order to determine the number of the regimes that best describe the observed data.

# 3 Simulations

## 3.1 Different parameter values across regimes

The following Data Generating Processes (DGP) are considered for the evaluation of the performance of the ICE algorithm in the detection of structural breaks. The following cases are examined:

1. DGP 1:Two Regimes with One Break

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 500 \\ 1 + 0.7x_{1t} + 0.5x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 501 \leq t \leq 1000 \end{cases} \tag{24}$$

2. DGP 2:Two Regimes with Two Breaks

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 330 \\ 1 + 0.7x_{1t} + 0.5x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 331 \leq t \leq 670 \\ 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 671 \leq t \leq 1000 \end{cases} \tag{25}$$

3. DGP 3:Three Regimes with Two Breaks

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 330 \\ 1 + 0.7x_{1t} + 0.5x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 331 \leq t \leq 670 \\ 1 + 0.2x_{1t} + 05x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 671 \leq t \leq 1000 \end{cases} \tag{26}$$

4. DGP 4:Four Regimes with Three Breaks

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 330 \\ 1 + 0.7x_{1t} + 0.5x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 331 \leq t \leq 670 \\ 1 + 0.2x_{1t} + 05x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 671 \leq t \leq 1000 \\ 1 + 1x_{1t} - 0.3x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 1001 \leq t \leq 1340 \end{cases} \tag{27}$$

The two regressors $X_1, X_2$ are random variables that are generated according to:

$$X_1 \sim N(1,1), \quad X_2 \sim B(2,1)$$

The choice of a linear regression model allows for a comparison of the proposed method's performance with the Bai and Perron method. The DGPs considered in the simulation studies are more general than those that exist in the literature in two respects. First, two regressors are considered instead of one, thus making the specification more realistic [9]. Second, specifications with more than one break are considered.

Before proceeding to discuss the simulation results, I need to emphasize the difference between identifying regimes and detecting breaks. As was discussed in previous sections, the ICE method is used to estimate regime switching models, and identify the location of the changing points through Bayesian segmentation. In this framework, structural breaks in the model parameters will occur at the regime switching points.

---

[9] Existing papers that consider simulation studies of the Bai and Perron method are restricted in using only one regressor that is generated according to a Normal distribution.

In most of the simulation specifications, the point where there is a regime switch coincides with a break point for the regression parameters. Regardless, this may not be the case in general as there can be cases where the process switches back to the original regime, as in the second DGP under consideration. Thus, whereas the Bai and Perron identifies break points, the IMI method needs to perform two tasks: a) it chooses the optimal number of regimes through the ICL-BIC criterion, and b) it identifies the location of regime switches. The regimes switching points for the ICE algorithm will correspond to the break points for the Bai and Perron method.

In view of these, the criteria for evaluating the Bai and Perron approach will be primarily whether it estimates the correct number of breaks or not and subsequently their locations, and then the quality of the regression parameter estimates. For the IMI case, the criteria will be whether it estimates correctly the number of different regimes, then the location of regime switches, and finally the quality of the parameter estimates within each regime.

A brief comparison of results obtained by the two approaches is presented in Tables 1 and 2, whereas the complete results from the simulations for all specifications, and for each methodology, are tabulated in 3 to 11. Overall, the results can be summarized as follows [10]:

The Bai and Perron method estimates the correct number of breaks in most of the replications, even though it tends to overestimate their number (Table 7). For DGP 1, it estimates the correct number of breaks (1) in 65% of the replications, and chooses two breaks 28% of the time. For DGPs 2 and 3, it estimates the correct number of breaks (2) in approximately 72 % of the replications, whereas for DGP 4 it estimates the correct number (3) in only 58% of the replications. The estimates of the break location are reported only for the cases when the BP methodology correctly identified the number of breaks (Table 8). It can be noticed that the median values for the locations either coincide or are very close to the true locations. Moreover, the estimates for the location of the breaks falls in the range of ±5 time intervals from the true value at 73.6 to 85.3 % of the replications that correctly identified the number of breaks.

The IMI method estimates the correct number of regimes quite accurately; in 100% of the replications for DGPs 1 and 2, in 93% for DGP 3, and in 74 % of the replications for DGP 4 (Table 4). The ICL-BIC criterion presents its minimum average value when the ICE algorithm is performed for the correct number of regimes (Table 3). It can also be noticed that as the number of regimes increases, their estimate becomes less accurate. In fact, there is a tendency to underfit the existing regimes. This is evident in the values of the ICL-BIC criterion; for DGP 4 there is an average difference of -14 between the criterion value for 4 and 3 regimes, whereas the difference between 4 and 5 regimes is approximately -100. This may be attributed to the nature of the ICE algorithm. Recall that the method simulates different state (regime) paths, thus effectively splitting the sample into the different regimes so as to increase the fit. If fewer regimes are assumed, this may provide a better fit especially when the parameters within each regime are such that the differences between the conditional distributions of the innovations are small.

In terms of the estimation of the regime switching points, which are comparable with the breaks points as estimated by the Bai and Perron methodology, the results are even more encouraging. When the correct number of regimes is picked by the ICL-BIC criterion, the correct number of regimes switches is always obtained and their descriptive statistics are those reported in this paper. The median values of switch locations are found to be one time interval further than the true values,and the estimates for the location of

---

[10] The results of Bai and Perron method correspond to a 5% significance level.

the switches falls in the range of ±5 time intervals from the true value at 85 to 93.9 % of the replications, when then number of regimes is correctly identified (Table 5). Moreover, even when the ICL-BIC criterion underestimates the number of the regimes, the IMI method still estimates the correct number of regime switches and their corresponding locations in most of the cases. For example, for DGP 4, when 3 regimes are picked, instead of 4, the correct value of the number of regimes switches (3) is estimated in 85% of the cases, with the corresponding switch locations falling within ±10 time intervals from the true values. In general, if the wrong number of 3 regimes were picked in all simulation replications, the 3 "correct" switches would be obtained in 79% of the cases, and at 7% of the cases 2 of the "correct" 3 switches[11].

Finally, in terms of the parameter estimates obtained by both approaches, Tables 6 and Tables 9 to 11 report estimates that correspond to the cases in which Bai and Perron pick the correct number of breaks, and the IMI method picks the correct number of regimes. It can be noticed that the estimates are quite accurate. This is expected; both approaches perform the estimation conditional on the knowledge of the locations of the breaks points (for Bai and Perron) or the number of regimes and the locations of the corresponding switches (for IMI). If the aforementioned estimates are close to the true values, then the parameter estimates should also be close to their true values.

Overall, the IMI method seems to perform better than the Bai and Perron method. It identifies the correct number of regime and the corresponding locations of regime switches more frequently, and even in the case that the ICL-BIC criterion suggests a smaller, than the true, value for the number of regimes, it can still identify correctly the true locations of regime switches.

## 3.2 Different models across regimes

The second part of the simulation study examines the performance of the ICE algorithm in the more interesting setting where different models are allowed to generate the observed data among different regimes.In particular, the following data generating processes (DGP) are considered:

1. Case I: Models 1,2

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 500 \\ 1 + 0.7x_{1t}x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.7^2\right), \quad 501 \leq t \leq 1000 \end{cases} \tag{28}$$

2. Case II: Models 1,3

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 500 \\ 0.5 + 0.7y_{t-1} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 501 \leq t \leq 1000 \end{cases} \tag{29}$$

3. Case II: Models 2,3

$$y_t = \begin{cases} 1 + 0.7x_{1t}x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.7^2\right), \quad 0 \leq t \leq 500 \\ 0.5 + 0.7y_{t-1} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 501 \leq t \leq 1000 \end{cases} \tag{30}$$

---

[11] The "correct" stands for a regime switching location estimated to be in the range of ±10 time intervals from the true value.

4. Case IV: Models 1,2,3

$$y_t = \begin{cases} 1 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.5^2\right), \quad 0 \leq t \leq 330 \\ 1 + 0.7x_{1t}x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.7^2\right), \quad 331 \leq t \leq 670 \\ 0.5 + 0.7x_{1t} - 0.25x_{2t} + \varepsilon_t & \varepsilon_t \sim N\left(0, 0.6^2\right), \quad 671 \leq t \leq 1000 \end{cases} \tag{31}$$

The regressors are generated, similarly to the previous section, according to:

$$X_1 \sim N(1,1), \quad X_2 \sim B(2,1).$$

It is apparent that the three types of models that may generate the data points within each regime are: 1) a linear regression with two regressors $X_1$, $X_2$; 2) a linear regression with the cross-product of the two variables $X_1 X_2$; and 3) an Autoregressive Model of order 1 (AR(1)). These are also the models that are considered in the IMI algorithm[12].

The criteria for the evaluation of the method will be similar to the case of different parameter values across regimes. It is of interest: a) whether the correct number of regimes will be picked by the ICL-BIC criterion, b) whether the correct types of models are obtained for each of the regimes, c) the estimates for the location of the regime switching points, and d) the quality of the estimates for the model parameters.

The simulation results are presented in Tables 12 to 15. The ICL-BIC criterion recognizes, on average, the correct number of regimes. Nevertheless, as in the previous simulation setting, there is a tendency to underfit by picking fewer regimes than actually exist, when the true number is large. For example, for Case IV, as Table 13 demonstrates, the ICL-BIC criterion picks the correct number of regimes (3) in 74% of the time but it also picks the lower, incorrect, number of regimes (2) in 26% of the replications. The difficulty in picking the true number of regimes is also apparent by the relative differences in the mean values, across replications, of the ICL-BIC criterion in Case IV. For example, the difference between three and two regimes is only -17, whereas the difference between three and four regimes is more approximately -140. Once more, this outcome can be attributed to the parameter values for the models that may lead to relatively close conditional distributions for the innovations across regimes. On the other hand, when only two regimes are considered, the accuracy reaches 100%.

The identification of the models is quite accurate; when the correct number of regimes is suggested by the ICL-BIC criterion, the types of models identified for the regimes are the ones that correspond to the true DGP. In terms of the locations of the regime switches, the results reported correspond to the cases that the true number of regimes are identified. The median estimates are only one time interval away from the true values, and the estimates for the location of the switches falls in the range of ±5 time intervals from the true value at 60.9 to 93.5 % of the replications. Similarly to the case of parameter changes across regimes, when a lower than the true number of regimes is picked, the locations of switches remain close to the "correct" values.

Finally, in terms of the parameter estimates, the tables report the values that correspond to the cases that the IMI method picks the correct number of regimes. It can be noticed that the estimates are quite accurate. This is expected since the number of regimes and the location of regime switches are accurately

---

[12] In economic applications, the types of models that may have generated the data according to each regime will be determined based on theoretical grounds. For example, regime switches in monetary policies could be analyzed using models that are variants of the Taylor rule.

estimated (see previous section's discussion).

Overall, the IMI method performs well in estimating all aspects of the simulations specifications, especially when the number of the regimes under considerations is small. Given that in most economic applications theoretical grounds suggest a small number of regimes, e.g. bull and bear stock markets, the ICE method appears to be a promising technique.

## 4 Conclusion

A complete procedure is introduced for the estimation of all the aspects of a Hidden Markov Model (HMM), in order to detect the existence of structural breaks and their corresponding locations. The structural breaks are defined as the data points where the underlying Markov Chain switches from one state to another.

The proposed procedure consists of three components. The first component is the estimation of the HMM with a variant of the Iterative Conditional Expectation-Generalized Mixture (ICE-GEMI) algorithm proposed by Delignon et al. (1997). The algorithm has a number of advantages over the traditional approach of estimating HMM using the Expectation Maximization (EM) algorithm: a) it is based on the notion of conditional expectation and not on the likelihood, thus it encompasses probability distributions that have both a discrete and a continuous part, which is a case where the notion of likelihood is no longer valid, b) it simulates regime paths in order to enable parameter estimation, which permits analysis of the conditional distributions of economic data, and allows for different functional forms across regimes.

The second component is the detection of the locations of regime switching by assigning states to data points according to the Maximum Posterior Mode (MPM) algorithm. The third, and final, component is the employment of the Integrated Classification Likelihood-Bayesian Information Criterion (ICL-BIC) for determining the number of regimes. This information criterion is essentially the BIC, but it also takes into account the classification of the data points to their corresponding regimes, and thus it is more suitable for use in the case of HMM.

The success of the overall procedure, denoted IMI by the initials of the component algorithms, to detect structural breaks is validated by two sets of simulations. In the first set of simulations only the parameters of a multiple linear regression model are permitted to differ across regimes. Due to the particular setting, the results are directly comparable with those provided by the Bai and Perron (1998) method. The IMI procedure fares better in all specifications, especially when the number of breaks is small. More importantly, even when the ICL-BIC component fails to correctly identify the number of regimes, the other two components can still provide correct estimates for the location of breaks points defined as regime switches. The second set of simulations permits differences in the functional forms of the conditional distribution of the dependent variable across regimes. The optimal model for each regime is decided using the BIC. The IMI procedure performed adequately, especially when the number of regimes is smaller. The procedure identified the correct number of regimes, the locations of the regime switching points, and the functional forms under each regime in the great majority of the replications.

Consequently, the IMI procedure is found to provide a useful overall approach to the estimation of all aspects related to a Hidden Markov Model, such as the number of regimes, the model parameters, and the location of regime switches. Nevertheless, it could be of interest to generalize the procedure by a) allowing multivariate variables for the dependent data, and b) examining nonparametric approaches in picking the correct model for each regime.

## References

[1] Akaike H., (1974). A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19, 716723.

[2] Alt R., W. Krämer, and W. Ploberger, (1988), Testing for Structural Change in Dynamic Models, *Econometrica*, 56, 1355-1369.

[3] Andrews D., (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica*, 61,821-856.

[4] Andrews D., and E. Zivot, (1992). Further Evidence on the Great Crash, the Oil-Price, and the Unit-Root Hypothesis, *Journal of Business and Economic Statistics*, 10, 251-270.

[5] Andrews D., and W. Ploberger, (1994). Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative, *Econometrica*, 62, 1383-1414.

[6] Bai J., (1994). Least Squares Estimation of a Shift in Linear Processes, *Journal of Time Series Analysis*, 15, 453-472.

[7] Bai J., (1997). Estimation of a Change Point in Multiple Regression Models, *Review of Economics and Statistics*, 79, 551-563.

[8] Bai J., and P.Perron, (1998). Estimating and testing linear models with multiple structural changes, *Econometrica*, 66, 4778.

[9] Bai J., and P.Perron, (2003). Computation and Analysis of Multiple Structural Change, *Journal of Applied Econometrics*, 18, 1-22.

[10] Baum L.E, T. Petrie, G. Soules and N. Weiss, (1970). A Maximization Technique in the Statistical Analysis of Probabilistic Functions of Finite State Markov Chains,*Annals of Mathematical Statistics*, 41, 164-171.

[11] Biernacki C., G. Celeux and G. Govaert, (1999). An Improvement of the NEC Criterion for Assessing the Number of Clusters in a Mixture Model, *Pattern Recognition Letters*, 20, 267-272.

[12] Biernacki C., G. Celeux, and G. Govaert, (2000). Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719-725.

[13] Bilmes J., (2006). What HMMs Can Do, *EICE Transactions on Information and Systems*, E89, 869-89.

[14] Brown, R.L., Durbin J., and Evans J.M., (1975). Techniques for Testing the Constancy of Regression Relationships Over Time, with Comments, *Journal of the Royal Statistical Society*, Ser. B, 37, 2,149-192.

[15] Chow G.C., (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions, *Econometrica*, 28, 591-605.

[16] Delignon Y., A. Marzouki, W. Pieczynski, (1997). Estimation of Generalized Mixture and its Application in Image Segmentation, *IEEE Transactions on Image Processing*, 6, 1364-1375.

[17] Delmas J.P., (1997). An Equivalence of the EM and ICE Algorithm for Exponential Family, *IEEE Transactions on Signal Processing*,45, 2613-2615.

[18] Dempster A. P., N.M Laird, and D.B. Rubin, (1977). Maximum Likelihood Estimation from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.

[19] Devijver P.A., and M. Dekesel, (1988). Champs aléatoires de Pickard et modelization d' images digitales, Traitement Du Signal, 5, 131-150.

[20] Engel C., (1994). Can the Markov Switching Model Forecast Exchange Rates?, *Journal of International Economics*, 36, 151-165.

[21] Fjortoft R., Y.Delignon, W.Pieczynski, M.Sigelle, and F. Tupin, (2003). Unsupervised Classification of Radar Images Using Hidden Markov Chains and Hidden Markov Random Fields, *IEEE Transactions on Geoscience and Remote Sensing*,41, 675- 686.

[22] Hamilton J. D., (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-384.

[23] Hansen B.E., (1997), Approximate Asymptotic p-values for Structural-change Tests, *Journal of Business and Economic Statistics*, 15, 60-67.

[24] Hansen B.E., (2001). The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity,*The Journal of Economic Perspectives*, 15,117-128.

[25] Hathaway R.J., (1986). A Constrained EM algorithm for univariate normal mixtures, *Journal of Statistical Computation and Simulation*, 23, 211230.

[26] Krämer W., and W. Ploberger, (1992), The CUSUM Test with OLS Residuals, *Econometrica*, 60, 271-285.

[27] Lumsdaine R., and D. Papell, (1997). Multiple Trend Breaks and the Unit-Root Hypothesis, *Review of Economics and Statistics*, 79, 212-218.

[28] McLachlan C. and D.Peel, (2000). *Finite Mixture Models*, New York: John Wiley and Sons.

[29] Marsh I.W., (2000). High-frequency Markov Switching Models in the Foreign Exchange Market, *Journal of Forecasting*, 19, 123-134.

[30] Perron P., (1989). The Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, *Econometrica*, 57, 1361-1401.

[31] Perron P., and Z. Qu, (2006). Estimating Restricted Structural Change Models, *Journal of Econometrics*, 134, 373-399.

[32] Pesaran M.H., and A. Timmermann, (2004). How Costly is it to Ignore Breaks When Forecasting the Direction of a Time Series?, *International Journal of Forecasting*, 20, 411-425.

[33] Pieczynski W., (1992). Statistical Image Segmentation, *Machine Graphics and Vision*, 1, 261-268.

[34] Pieczynski W., (1994). Champs de Markov Cachés et Estimation Conditionelle Itérative, *Traitement du Signal*, 11, 141-153.

[35] Quandt R., (1960). Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes, *Journal of the American Statistical Association*, 55, 324-330.

[36] Rabiner L.R., (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceedings of the IEEE*, 77, 257-286.

[37] Resnick S.I, (1992).*Adventures in Stochastic Processes*, Boston: Birkhäuser.

[38] Schaller H., and S. van Norden, (1997). Regime Switching in Stock Market Returns, *Applied Financial Economics*, 7, 177-191.

[39] Schwarz, G., (1978). Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461-464.

# A Tables

## A.1 Simulation Results: Different parameters across regimes

### A.1.1 IMI vs B&P summary results

| **IMI vs B&P: No of Regimes/Breaks Frequencies** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **2 Regimes - 1 Break** | | **2 Regimes - 2 Breaks** | | **3 Regimes - 2 Breaks** | | **4 Regimes - 3 Breaks** | |
| **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** |
| **IMI: No of Regimes Identified** | | | | | | | |
| 1 | 0.0 | 1 | 0.0 | 1 | 0.0 | 1 | 0.0 |
| 2* | 100.0 | 2* | 100.0 | 2 | 3.0 | 2 | 1.0 |
| 3 | 0.0 | 3 | 0.0 | 3* | 93.0 | 3 | 24.0 |
| 4 | 0.0 | 4 | 0.0 | 4 | 4.0 | 4* | 74.0 |
| 5 | 0.0 | 5 | 0.0 | 5 | 0.0 | 5 | 1.0 |
| | | | | | | 6 | 0.0 |
| *: true number of regimes | | | | | | | |
| **B&P: No of Breaks Identified** | | | | | | | |
| 1* | 65.0 | 1 | 0.0 | 1 | 0.0 | 1 | 0.0 |
| 2 | 28.0 | 2* | 72.0 | 2* | 71.5 | 2 | 0.0 |
| 3 | 6.5 | 3 | 25.5 | 3 | 27.0 | 3* | 58.0 |
| 4 | 0.5 | 4 | 2.5 | 4 | 1.5 | 4 | 37.0 |
| 5 | 0.0 | 5 | 0.0 | 5 | 0.0 | 5 | 6.0 |
| | | | | | | 6 | 0.0 |
| *: true number of breaks | | | | | | | |

Table 1: IMI vs B&P summary results, Part 1

| **Break Locations** | | | | | | |
|---|---|---|---|---|---|---|
| | **B&P Sequential** | | | **IMI** | | |
| | **4 Regimes -3 Breaks** | | | | | |
| **True Value** | **330** | **670** | **1000** | **330** | **670** | **1000** |
| **mean** | 329.4 | 672.0 | 1001.9 | 331.06 | 670.6 | 1000.9 |
| **median** | 330 | 670 | 1000 | 331 | 671 | 1001 |
| **b-5<%<b+5** | 84.5 | 77.6 | 85.3 | 84.5 | 87.8 | 93.9 |
| | **3 Regimes - 2 Breaks** | | | | | |
| **True Value** | **330** | **670** | | **330** | **670** | |
| **mean** | 329.5 | 672.1 | | 330.7 | 671.3 | |
| **median** | 330 | 671 | | 331 | 671 | |
| **b-5<%<b+5** | 84.6 | 80.4 | | 90.0 | 92.1 | |
| | **2 Regimes - 2 Breaks** | | | | | |
| **True Value** | **330** | **670** | | **330** | **670** | |
| **mean** | 327.0 | 672.6 | | 330.6 | 671.1 | |
| **median** | 329 | 670 | | 331 | 671 | |
| **b-5<%<b+5** | 73.6 | 77.1 | | 91.5 | 86.9 | |
| | **2 Regimes - 1 Break** | | | | | |
| **True Value** | **500** | | | **500** | | |
| **mean** | 500 | | | 501.2 | | |
| **median** | 500 | | | 501 | | |
| **b-5<%<b+5** | 83.1 | | | 85 | | |

Table 2: IMI vs B&P summary results, Part 2

### A.1.2 ICE Methodology

| **ICL-BIC Criterion** | | | | | | |
|---|---|---|---|---|---|---|
| | **4 Regimes -3 Breaks** | | | | | |
| **no Regimes** | **1** | **2** | **3** | **4*** | **5** | **6** |
| **mean** | 2721.07 | 2573.42 | 2502.07 | 2486.02 | 2610.24 | 2722.83 |
| **median** | 2719.50 | 2569.27 | 2502.99 | 2476.00 | 2575.87 | 2686.52 |
| **stdev** | 48.92 | 64.28 | 65.48 | 93.51 | 123.84 | 133.56 |
| **min** | 2586.87 | 2447.76 | 2343.62 | 2310.13 | 2403.55 | 2511.68 |
| **max** | 2856.90 | 2758.26 | 2690.79 | 3247.90 | 3029.72 | 3146.49 |
| | **3 Regimes -2 Breaks** | | | | | |
| **no Regimes** | **1** | **2** | **3*** | **4** | **5** | **6** |
| **mean** | 1957.01 | 1828.14 | 1746.26 | 1873.33 | 1985.09 | |
| **median** | 1955.26 | 1829.14 | 1742.22 | 1834.14 | 1932.18 | |
| **stdev** | 45.97 | 47.71 | 55.91 | 122.05 | 145.03 | |
| **min** | 1841.02 | 1709.34 | 1616.43 | 1692.42 | 1782.23 | |
| **max** | 2060.46 | 1939.10 | 1962.76 | 2260.50 | 2475.58 | |
| | **2 Regimes -2 Breaks** | | | | | |
| **no Regimes** | **1** | **2*** | **3** | **4** | **5** | **6** |
| **mean** | 1698.57 | 1567.61 | 1979.83 | 2158.08 | | |
| **median** | 1704.03 | 1572.24 | 1936.12 | 2125.52 | | |
| **stdev** | 44.10 | 47.11 | 309.03 | 371.99 | | |
| **min** | 1580.49 | 1455.24 | 1529.20 | 1601.42 | | |
| **max** | 1827.51 | 1698.49 | 2530.14 | 3022.35 | | |
| | **2 Regimes -1 Break** | | | | | |
| **no Regimes** | **1** | **2*** | **3** | **4** | **5** | **6** |
| **mean** | 1724.35 | 1543.21 | 1828.94 | 1952.79 | | |
| **median** | 1724.56 | 1545.18 | 1795.38 | 1904.65 | | |
| **stdev** | 46.28 | 45.92 | 227.04 | 270.18 | | |
| **min** | 1568.57 | 1406.84 | 1492.16 | 1566.51 | | |
| **max** | 1845.14 | 1652.38 | 2268.93 | 2828.84 | | |
| | *true value | | | | | |

Table 3: IMI: ICL-BIC Criterion

| **No of Regimes Identified** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **2 Regimes - 1 Break** | | **2 Regimes - 2 Breaks** | | **3 Regimes - 2 Breaks** | | **4 Regimes - 3 Breaks** | |
| **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** |
| **1** | 0.0 | **1** | 0.0 | **1** | 0.0 | **1** | 0.0 |
| **2*** | 100.0 | **2*** | 100.0 | **2** | 3.0 | **2** | 1.0 |
| **3** | 0.0 | **3** | 0.0 | **3*** | 93.0 | **3** | 24.0 |
| **4** | 0.0 | **4** | 0.0 | **4** | 4.0 | **4*** | 74.0 |
| **5** | 0.0 | **5** | 0.0 | **5** | 0.0 | **5** | 1.0 |
| | | | | | | **6** | 0.0 |
| *: true number of regimes | | | | | | | |

Table 4: IMI: No of Regimes Identified

| **Break Locations** | | | | | |
|---|---|---|---|---|---|
| | **4 Regimes -3 Breaks** | | | **3 Regimes -2 Breaks** | |
| **True Locations** | **330** | **670** | **1000** | **330** | **670** |
| **mean** | 331.06 | 670.56 | 1000.94 | 330.65 | 671.26 |
| **median** | 331 | 671 | 1001 | 331 | 671 |
| **stdev** | 5.15 | 4.03 | 3.24 | 3.93 | 3.13 |
| **min** | 308 | 651 | 981 | 303 | 657 |
| **max** | 363 | 685 | 1019 | 341 | 684 |
| **b-5<%<b+5** | 84.5 | 87.8 | 93.9 | 90.0 | 92.1 |
| | **2 Regimes -2 Breaks** | | | **2 Regimes -1 Break** | |
| **True Locations** | **330** | **670** | | **500** | |
| **mean** | 330.62 | 671.11 | | 501.22 | |
| **median** | 331 | 671 | | 501 | |
| **stdev** | 3.95 | 3.91 | | 4.02 | |
| **min** | 300 | 652 | | 485 | |
| **max** | 345 | 686 | | 513 | |
| **b-5<%<b+5** | 91.5 | 86.9 | | 85.0 | |

Table 5: IMI: Break Locations

| **Parameter Estimates** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Case I: 2 Regimes - 1 Break** | | | | | **Case II: 2 Regimes - 2 Breaks** | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Regime 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | **-0.25** | **0.5** | **True Values** | **1** | **0.7** | **-0.25** | **0.5** |
| **mean** | 0.997 | 0.700 | -0.243 | 0.500 | **mean** | 1.003 | 0.700 | -0.255 | 0.498 |
| **median** | 0.996 | 0.701 | -0.246 | 0.500 | **median** | 0.994 | 0.701 | -0.253 | 0.499 |
| **stdev** | 0.073 | 0.023 | 0.094 | 0.015 | **stdev** | 0.063 | 0.021 | 0.080 | 0.014 |
| **min** | 0.777 | 0.639 | -0.471 | 0.462 | **min** | 0.814 | 0.626 | -0.450 | 0.459 |
| **max** | 1.174 | 0.775 | 0.072 | 0.539 | **max** | 1.163 | 0.745 | -0.046 | 0.540 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Regime 2** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | **0.5** | **0.5** | **True Values** | **1** | **0.7** | **0.5** | **0.5** |
| **mean** | 0.996 | 0.702 | 0.507 | 0.499 | **mean** | 0.992 | 0.706 | 0.504 | 0.500 |
| **median** | 0.999 | 0.702 | 0.507 | 0.498 | **median** | 0.988 | 0.704 | 0.510 | 0.500 |
| **stdev** | 0.067 | 0.022 | 0.086 | 0.016 | **stdev** | 0.091 | 0.028 | 0.126 | 0.018 |
| **min** | 0.800 | 0.647 | 0.302 | 0.462 | **min** | 0.780 | 0.636 | 0.179 | 0.453 |
| **max** | 1.144 | 0.762 | 0.758 | 0.539 | **max** | 1.199 | 0.784 | 0.787 | 0.558 |
| **Case III: 3 Regimes - 2 Breaks** | | | | | **Case IV: 4 Regimes - 3 Breaks** | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Regime 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | **-0.25** | **0.5** | **True Values** | **1** | **0.7** | **-0.25** | **0.5** |
| **mean** | 0.999 | 0.700 | -0.250 | 0.501 | **mean** | 1.002 | 0.700 | -0.251 | 0.500 |
| **median** | 1.006 | 0.703 | -0.259 | 0.501 | **median** | 1.003 | 0.696 | -0.252 | 0.501 |
| **stdev** | 0.083 | 0.038 | 0.114 | 0.021 | **stdev** | 0.082 | 0.030 | 0.101 | 0.021 |
| **min** | 0.797 | 0.434 | -0.694 | 0.437 | **min** | 0.761 | 0.605 | -0.484 | 0.434 |
| **max** | 1.284 | 0.774 | -0.002 | 0.611 | **max** | 1.208 | 0.848 | 0.004 | 0.554 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Regime 2** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | **0.5** | **0.5** | **True Values** | **1** | **0.7** | **0.5** | **0.5** |
| **mean** | 1.002 | 0.702 | 0.494 | 0.500 | **mean** | 1.004 | 0.705 | 0.478 | 0.502 |
| **median** | 1.003 | 0.701 | 0.495 | 0.500 | **median** | 1.000 | 0.701 | 0.485 | 0.499 |
| **stdev** | 0.087 | 0.028 | 0.116 | 0.018 | **stdev** | 0.091 | 0.045 | 0.133 | 0.028 |
| **min** | 0.769 | 0.632 | 0.158 | 0.456 | **min** | 0.805 | 0.462 | 0.055 | 0.418 |
| **max** | 1.245 | 0.796 | 0.830 | 0.542 | **max** | 1.392 | 0.875 | 0.809 | 0.614 |
| **Regime 3** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Regime 3** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.2** | **0.5** | **0.6** | **True Values** | **1** | **0.2** | **0.5** | **0.6** |
| **mean** | 1.009 | 0.214 | 0.464 | 0.595 | **mean** | 0.998 | 0.200 | 0.505 | 0.598 |
| **median** | 1.008 | 0.199 | 0.480 | 0.597 | **median** | 1.001 | 0.195 | 0.503 | 0.597 |
| **stdev** | 0.108 | 0.085 | 0.183 | 0.024 | **stdev** | 0.102 | 0.032 | 0.143 | 0.023 |
| **min** | 0.660 | 0.092 | -0.275 | 0.530 | **min** | 0.763 | 0.121 | 0.181 | 0.534 |
| **max** | 1.278 | 0.707 | 0.889 | 0.660 | **max** | 1.226 | 0.285 | 0.838 | 0.646 |
| | | | | | **Regime 4** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| | | | | | **True Values** | **1** | **1** | **-0.3** | **0.6** |
| | | | | | **mean** | 1.007 | 0.986 | -0.312 | 0.600 |
| | | | | | **median** | 1.006 | 0.998 | -0.323 | 0.599 |
| | | | | | **stdev** | 0.105 | 0.057 | 0.139 | 0.024 |
| | | | | | **min** | 0.694 | 0.807 | -0.648 | 0.532 |
| | | | | | **max** | 1.220 | 1.085 | 0.166 | 0.654 |

Table 6: IMI: Parameter Estimates

### A.1.3 Bai & Perron Methodology

| **No of Breaks Identified** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **2 Regimes - 1 Break** | | **2 Regimes - 2 Breaks** | | **3 Regimes - 2 Breaks** | | **4 Regimes - 3 Breaks** | |
| **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** |
| **1*** | 65.0 | **1** | 0.0 | **1** | 0.0 | **1** | 0.0 |
| **2** | 28.0 | **2*** | 72.0 | **2*** | 71.5 | **2** | 0.0 |
| **3** | 6.5 | **3** | 25.5 | **3** | 27.0 | **3*** | 58.0 |
| **4** | 0.5 | **4** | 2.5 | **4** | 1.5 | **4** | 37.0 |
| **5** | 0.0 | **5** | 0.0 | **5** | 0.0 | **5** | 6.0 |
| | | | | | | **6** | 0.0 |
| *: true number of breaks | | | | | | | |

Table 7: B&P: No of Breaks Identified

| **Break Locations** | | | | | | |
|---|---|---|---|---|---|---|
| | **Sequential** | | | **Repartition** | | |
| | **4 Regimes -3 Breaks** | | | | | |
| **True Value** | **330** | **670** | **1000** | **330** | **670** | **1000** |
| **mean** | 329.4 | 672.0 | 1001.9 | 329.4 | 672.0 | 1001.9 |
| **median** | 330 | 670 | 1000 | 330 | 670 | 1000 |
| **stdev** | 5.9 | 5.9 | 4.4 | 5.9 | 5.9 | 4.4 |
| **min** | 304 | 659 | 993 | 304 | 659 | 993 |
| **max** | 356 | 700 | 1020 | 356 | 700 | 1020 |
| **b-5<%<b+5** | 84.5 | 77.6 | 85.3 | 84.5 | 77.6 | 85.3 |
| | **3 Regimes - 2 Breaks** | | | | | |
| **True Value** | **330** | **670** | | **330** | **670** | |
| **mean** | 329.5 | 672.1 | | 329.5 | 672.1 | |
| **median** | 330 | 671 | | 330 | 671 | |
| **stdev** | 4.6 | 5.4 | | 4.6 | 5.4 | |
| **min** | 311 | 660 | | 311 | 660 | |
| **max** | 351 | 707 | | 351 | 707 | |
| **b-5<%<b+5** | 84.6 | 80.4 | | 84.0 | 79.9 | |
| | **2 Regimes - 2 Breaks** | | | | | |
| **True Value** | **330** | **670** | | **330** | **670** | |
| **mean** | 327.0 | 672.6 | | 327.0 | 672.6 | |
| **median** | 329 | 670 | | 329 | 670 | |
| **stdev** | 7.13 | 7.88 | | 7.13 | 7.88 | |
| **min** | 289 | 651 | | 289 | 651 | |
| **max** | 342 | 709 | | 342 | 709 | |
| **b-5<%<b+5** | 73.6 | 77.1 | | 73.6 | 77.1 | |
| | **2 Regimes - 1 Break** | | | | | |
| **True Value** | **500** | | | **500** | | |
| **mean** | 500 | | | 500 | | |
| **median** | 500 | | | 500 | | |
| **stdev** | 5 | | | 5 | | |
| **min** | 476 | | | 476 | | |
| **max** | 515 | | | 515 | | |
| **b-5<%<b+5** | 83.1 | | | 83.1 | | |

Table 8: B&P: Break Locations

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| | **Sequential** | | | **Repartition** | | |
| | **2 Regimes - 1 Break** | | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **-0.25** | **1** | **0.7** | **-0.25** |
| **mean** | 0.9978 | 0.7013 | -0.2547 | 0.9978 | 0.7013 | -0.2547 |
| **median** | 1.0029 | 0.7024 | -0.2485 | 1.0029 | 0.7024 | -0.2485 |
| **stdev** | 0.0713 | 0.0212 | 0.1005 | 0.0713 | 0.0212 | 0.1005 |
| **min** | 0.7733 | 0.6218 | -0.5468 | 0.7733 | 0.6218 | -0.5468 |
| **max** | 1.1895 | 0.7527 | 0.0475 | 1.1895 | 0.7527 | 0.0475 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **0.5** | **1** | **0.7** | **0.5** |
| **mean** | 1.0062 | 0.7008 | 0.4935 | 1.0062 | 0.7008 | 0.4935 |
| **median** | 1.0114 | 0.7015 | 0.4850 | 1.0114 | 0.7015 | 0.4850 |
| **stdev** | 0.0709 | 0.0219 | 0.1021 | 0.0709 | 0.0219 | 0.1021 |
| **min** | 0.7985 | 0.6396 | 0.2574 | 0.7985 | 0.6396 | 0.2574 |
| **max** | 1.1559 | 0.7611 | 0.7916 | 1.1559 | 0.7611 | 0.7916 |
| | **2 Regimes - 2 Breaks** | | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **-0.25** | **1** | **0.7** | **-0.25** |
| **mean** | 1.0019 | 0.7001 | -0.2585 | 1.0025 | 0.7001 | -0.2617 |
| **median** | 1.0001 | 0.7012 | -0.2494 | 1.0001 | 0.7012 | -0.2494 |
| **stdev** | 0.0866 | 0.0289 | 0.1173 | 0.0868 | 0.0288 | 0.1181 |
| **min** | 0.8221 | 0.6200 | -0.5471 | 0.8221 | 0.6200 | -0.5642 |
| **max** | 1.2556 | 0.7650 | -0.0087 | 1.2556 | 0.7609 | -0.0087 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **0.5** | **1** | **0.7** | **0.5** |
| **mean** | 1.0115 | 0.6995 | 0.4889 | 1.0102 | 0.6997 | 0.4829 |
| **median** | 1.0072 | 0.6989 | 0.5016 | 1.0072 | 0.6976 | 0.4989 |
| **stdev** | 0.0933 | 0.0266 | 0.1214 | 0.0935 | 0.0263 | 0.1218 |
| **min** | 0.7826 | 0.6379 | 0.1594 | 0.7826 | 0.6397 | 0.1803 |
| **max** | 1.2739 | 0.7602 | 0.7891 | 1.2739 | 0.7679 | 0.7891 |
| **Regime 1a** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **-0.25** | **1** | **0.7** | **-0.25** |
| **mean** | 0.9945 | 0.7032 | -0.2503 | 0.9947 | 0.7027 | -0.2522 |
| **median** | 0.9823 | 0.7036 | -0.2427 | 0.9835 | 0.7033 | -0.2462 |
| **stdev** | 0.0883 | 0.0249 | 0.1293 | 0.0890 | 0.0253 | 0.1310 |
| **min** | 0.7862 | 0.6428 | -0.6139 | 0.7862 | 0.6354 | -0.6139 |
| **max** | 1.2812 | 0.7565 | 0.0759 | 1.2812 | 0.7565 | 0.0759 |

Table 9: B&P: Parameter Estimates, Part 1

| **Parameter Estimates** | | | | | | |
|---|---|---|---|---|---|---|
| | **Sequential** | | | **Repartition** | | |
| | **3 Regimes - 2 Breaks** | | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **-0.25** | **1** | **0.7** | **-0.25** |
| **mean** | 1.0012 | 0.7009 | -0.2562 | 1.0012 | 0.7009 | -0.2562 |
| **median** | 0.9964 | 0.7016 | -0.2483 | 0.9964 | 0.7016 | -0.2483 |
| **stdev** | 0.0879 | 0.0281 | 0.1197 | 0.0879 | 0.0281 | 0.1197 |
| **min** | 0.8221 | 0.6200 | -0.5471 | 0.8221 | 0.6200 | -0.5471 |
| **max** | 1.2556 | 0.7650 | -0.0087 | 1.2556 | 0.7650 | -0.0087 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **0.5** | **1** | **0.7** | **0.5** |
| **mean** | 1.0099 | 0.6997 | 0.4839 | 1.0097 | 0.6996 | 0.4888 |
| **median** | 1.0061 | 0.6987 | 0.4950 | 1.0084 | 0.6997 | 0.4969 |
| **stdev** | 0.0922 | 0.0261 | 0.1201 | 0.0925 | 0.0260 | 0.1206 |
| **min** | 0.7871 | 0.6415 | 0.1719 | 0.7889 | 0.6386 | 0.1719 |
| **max** | 1.2690 | 0.7630 | 0.7838 | 1.2705 | 0.7630 | 0.7823 |
| **Regime 3** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.2** | **0.5** | **1** | **0.2** | **0.5** |
| **mean** | 0.9929 | 0.2016 | 0.5055 | 0.9922 | 0.2037 | 0.5024 |
| **median** | 0.9846 | 0.2039 | 0.5085 | 0.9825 | 0.2045 | 0.5045 |
| **stdev** | 0.1028 | 0.0289 | 0.1533 | 0.1023 | 0.0287 | 0.1512 |
| **min** | 0.7638 | 0.1306 | 0.0696 | 0.7512 | 0.1349 | 0.0696 |
| **max** | 1.3351 | 0.2678 | 0.8753 | 1.3351 | 0.2678 | 0.8753 |

Table 10: B&P: Parameter Estimates, Part 2

| **Parameter Estimates** | | | | | | |
|---|---|---|---|---|---|---|
| | **Sequential** | | | **Repartition** | | |
| | **4 Regimes - 3 Breaks** | | | | | |
| **Regime 1** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **-0.25** | **1** | **0.7** | **-0.25** |
| **mean** | 0.9984 | 0.7016 | -0.2502 | 0.9984 | 0.7016 | -0.2502 |
| **median** | 0.9975 | 0.7025 | -0.2568 | 0.9975 | 0.7025 | -0.2568 |
| **stdev** | 0.0873 | 0.0310 | 0.1166 | 0.0873 | 0.0310 | 0.1166 |
| **min** | 0.6930 | 0.6333 | -0.5358 | 0.6930 | 0.6333 | -0.5358 |
| **max** | 1.1827 | 0.7634 | 0.1060 | 1.1827 | 0.7634 | 0.1060 |
| **Regime 2** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.7** | **0.5** | **1** | **0.7** | **0.5** |
| **mean** | 1.0078 | 0.6997 | 0.4855 | 1.0078 | 0.7001 | 0.4886 |
| **median** | 1.0145 | 0.7009 | 0.4735 | 1.0145 | 0.7001 | 0.4830 |
| **stdev** | 0.0839 | 0.0278 | 0.1210 | 0.0838 | 0.0278 | 0.1204 |
| **min** | 0.7074 | 0.6143 | 0.2289 | 0.7084 | 0.6143 | 0.2289 |
| **max** | 1.1877 | 0.7669 | 0.8543 | 1.1877 | 0.7669 | 0.8450 |
| **Regime 3** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **0.2** | **0.5** | **1** | **0.2** | **0.5** |
| **mean** | 0.9962 | 0.2039 | 0.4929 | 0.9974 | 0.2025 | 0.4920 |
| **median** | 0.9975 | 0.2029 | 0.4812 | 0.9995 | 0.1979 | 0.4747 |
| **stdev** | 0.1011 | 0.0338 | 0.1304 | 0.1029 | 0.0322 | 0.1319 |
| **min** | 0.7190 | 0.1271 | 0.0814 | 0.7255 | 0.1273 | 0.0825 |
| **max** | 1.2889 | 0.2999 | 0.8228 | 1.2946 | 0.2922 | 0.8210 |
| **Regime 4** | **ct** | $X_1$ | $X_2$ | **ct** | $X_1$ | $X_2$ |
| **True values** | **1** | **1** | **-0.3** | **1** | **1** | **-0.3** |
| **mean** | 0.9860 | 1.0036 | -0.2833 | 0.9849 | 1.0020 | -0.2796 |
| **median** | 0.9788 | 1.0029 | -0.2630 | 0.9769 | 0.9996 | -0.2630 |
| **stdev** | 0.0898 | 0.0321 | 0.1212 | 0.0887 | 0.0317 | 0.1196 |
| **min** | 0.8178 | 0.9174 | -0.5915 | 0.7904 | 0.9148 | -0.5915 |
| **max** | 1.2807 | 1.0932 | -0.0344 | 1.2807 | 1.0884 | -0.0344 |

Table 11: B&P: Parameter Estimates, Part 3

## A.2 Simulation Results: Different models across regimes

| **ICL-BIC Criterion** | | | | | |
|---|---|---|---|---|---|
| | **Models 1,2,3** | | | | |
| **no Regimes** | **1** | **2** | **3*** | **4** | **5** |
| **mean** | 2331.73 | 2002.92 | 1983.76 | 2126.51 | 2248.06 |
| **median** | 2336.60 | 2000.78 | 1985.07 | 2088.86 | 2190.88 |
| **stdev** | 64.38 | 59.30 | 50.63 | 130.39 | 147.20 |
| **min** | 2163.91 | 1831.89 | 1820.31 | 1896.23 | 1986.70 |
| **max** | 2503.81 | 2247.76 | 2180.76 | 2485.47 | 2675.58 |
| | **Models 1,2** | | | | |
| **no Regimes** | **1** | **2*** | **3** | **4** | **5** |
| **mean** | 1971.69 | 1892.05 | 2183.44 | 2324.44 | |
| **median** | 1975.12 | 1888.46 | 2165.31 | 2260.08 | |
| **stdev** | 48.38 | 47.11 | 235.99 | 314.11 | |
| **min** | 1847.51 | 1753.54 | 1811.42 | 1887.34 | |
| **max** | 2089.22 | 2009.79 | 2640.28 | 3814.07 | |
| | **Models 1,3** | | | | |
| **no Regimes** | **1** | **2*** | **3** | **4** | **5** |
| **mean** | 2352.38 | 1723.80 | 2020.75 | 2161.93 | |
| **median** | 2355.09 | 1725.08 | 1997.79 | 2141.07 | |
| **stdev** | 66.45 | 43.28 | 240.30 | 289.08 | |
| **min** | 2174.45 | 1568.76 | 1672.19 | 1756.87 | |
| **max** | 2530.84 | 1840.20 | 2476.51 | 2933.83 | |
| | **Models 2,3** | | | | |
| **no Regimes** | **1** | **2*** | **3** | **4** | **5** |
| **mean** | 2430.83 | 2066.82 | 2390.82 | 2487.24 | |
| **median** | 2434.88 | 2067.58 | 2397.21 | 2425.48 | |
| **stdev** | 53.46 | 43.76 | 232.39 | 281.55 | |
| **min** | 2250.01 | 1919.79 | 2031.49 | 2057.94 | |
| **max** | 2568.53 | 2195.68 | 2847.39 | 3363.24 | |
| | *true value | | | | |

Table 12: ICL-BIC Criterion

| **No of Regimes Identified** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Models 1,2,3** | | **Models 1,2** | | **Models 1,3** | | **Models 2,3** | |
| **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** | **No** | **% of Sim** |
| **1** | 0.0 | **1** | 0.0 | **1** | 0.0 | **1** | 0.0 |
| **2** | 26.0 | **2*** | 98.0 | **2*** | 100.0 | **2*** | 100.0 |
| **3*** | 74.0 | **3** | 2.0 | **3** | 0.0 | **3** | 0.0 |
| **4** | 0.0 | **4** | 0.0 | **4** | 0.0 | **4** | 0.0 |
| **5** | 0.0 | | | | | | |
| *: true number of regimes | | | | | | | |

Table 13: No of Regimes Identified

| **Break Locations** | | | |
|---|---|---|---|
| **Models** | **1,2,3** | | |
| **True Locations** | **330** | **670** | |
| **mean** | 331.89 | 670.66 | |
| **median** | 331 | 671 | |
| **stdev** | 8.46 | 4.58 | |
| **min** | 281 | 643 | |
| **max** | 371 | 689 | |
| **b-5<%<b+5** | 70.1 | 89.2 | |
| **Models** | **1,2** | **1,3** | **2,3** |
| **True Locations** | **500** | **500** | **500** |
| **mean** | 500.49 | 501.05 | 500.96 |
| **median** | 501 | 501 | 501 |
| **stdev** | 12.24 | 2.60 | 4.43 |
| **min** | 422 | 492 | 479 |
| **max** | 552 | 513 | 524 |
| **b-5<%<b+5** | 60.9 | 93.5 | 86.0 |

Table 14: Break Locations

| **Parameter Estimates** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Case I: Models 1, 2** | | | | | **Case II: Models 1, 3** | | | | |
| **Model 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ | **Model 1** | **ct** | $X_1$ | $X_1$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | **-0.25** | **0.5** | **True Values** | **1** | **0.7** | **-0.25** | **0.5** |
| **mean** | 1.001 | 0.701 | -0.251 | 0.500 | **mean** | 0.998 | 0.699 | -0.247 | 0.498 |
| **median** | 0.999 | 0.701 | -0.254 | 0.499 | **median** | 1.000 | 0.698 | -0.248 | 0.498 |
| **stdev** | 0.073 | 0.024 | 0.092 | 0.016 | **stdev** | 0.069 | 0.022 | 0.096 | 0.016 |
| **min** | 0.813 | 0.634 | -0.518 | 0.465 | **min** | 0.804 | 0.641 | -0.490 | 0.449 |
| **max** | 1.192 | 0.765 | -0.017 | 0.548 | **max** | 1.161 | 0.767 | 0.022 | 0.537 |
| **Model 2** | **ct** | $X_1X_2$ | | $\sigma_\epsilon$ | **Model 3** | **ct** | $Y_{t-1}$ | | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | | **0.7** | **True Values** | **0.5** | **0.7** | | **0.6** |
| **mean** | 1.000 | 0.698 | | 0.699 | **mean** | 0.512 | 0.693 | | 0.602 |
| **median** | 1.001 | 0.699 | | 0.699 | **median** | 0.506 | 0.696 | | 0.602 |
| **stdev** | 0.041 | 0.039 | | 0.024 | **stdev** | 0.063 | 0.032 | | 0.018 |
| **min** | 0.891 | 0.596 | | 0.638 | **min** | 0.364 | 0.600 | | 0.536 |
| **max** | 1.095 | 0.800 | | 0.757 | **max** | 0.683 | 0.771 | | 0.644 |
| **Case III: Models 2, 3** | | | | | **Case IV: Models 1, 2, 3** | | | | |
| **Model 2** | **ct** | $X_1X_2$ | | $\sigma_\epsilon$ | **Model 1** | **ct** | $X_1$ | $X_2$ | $\sigma_\epsilon$ |
| **True Values** | **1** | **0.7** | | **0.7** | **True Values** | **1** | **0.7** | **-0.25** | **0.5** |
| **mean** | 1.002 | 0.698 | | 0.701 | **mean** | 1.006 | 0.698 | -0.257 | 0.499 |
| **median** | 1.001 | 0.695 | | 0.701 | **median** | 1.016 | 0.698 | -0.251 | 0.501 |
| **stdev** | 0.041 | 0.044 | | 0.020 | **stdev** | 0.086 | 0.027 | 0.120 | 0.020 |
| **min** | 0.890 | 0.543 | | 0.647 | **min** | 0.803 | 0.627 | -0.572 | 0.446 |
| **max** | 1.124 | 0.827 | | 0.752 | **max** | 1.216 | 0.770 | 0.023 | 0.574 |
| **Model 3** | **ct** | $Y_{t-1}$ | | $\sigma_\epsilon$ | **Model 2** | **ct** | $X_1X_2$ | | $\sigma_\epsilon$ |
| **True Values** | **0.5** | **0.7** | | **0.6** | **True Values** | **1** | **0.7** | | **0.7** |
| **mean** | 0.510 | 0.694 | | 0.600 | **mean** | 0.998 | 0.697 | | 0.701 |
| **median** | 0.507 | 0.695 | | 0.599 | **median** | 1.000 | 0.693 | | 0.703 |
| **stdev** | 0.058 | 0.030 | | 0.021 | **stdev** | 0.049 | 0.053 | | 0.028 |
| **min** | 0.379 | 0.602 | | 0.524 | **min** | 0.856 | 0.519 | | 0.620 |
| **max** | 0.658 | 0.766 | | 0.669 | **max** | 1.112 | 0.861 | | 0.765 |
| | | | | | **Model 3** | **ct** | $Y_{t-1}$ | | $\sigma_\epsilon$ |
| | | | | | **True Values** | **0.5** | **0.7** | | **0.6** |
| | | | | | **mean** | 0.520 | 0.690 | | 0.599 |
| | | | | | **median** | 0.518 | 0.694 | | 0.597 |
| | | | | | **stdev** | 0.079 | 0.042 | | 0.025 |
| | | | | | **min** | 0.313 | 0.567 | | 0.547 |
| | | | | | **max** | 0.726 | 0.784 | | 0.660 |

Table 15: Parameter Estimates